11-16-00

A

JC715 U.S. PTO 11/15/00

# BAKER BOTTS L.L.P.

30 ROCKEFELLER PLAZA
44TH FLOOR
NEW YORK, NEW YORK
10112-4498
212 705 5000
FAX 212 705 5020

AUSTIN
BAKU
DALLAS
HOUSTON
LONDON
MOSCOW
**NEW YORK**
WASHINGTON

Appln. Trans.
PATENT

JC658 U.S. PTO 09/713794 11/15/00

**UTILITY PATENT APPLICATION TRANSMITTAL**
*(Only for new nonprovisional applications under 37 CFR 1.53(b))*

Attorney Docket No. A32000-A-072667.0172

First Named Inventor YANNICK BATARD

Express Mail Label No. EK938097140US

Total Pages ___

November 15, 2000

**BY EXPRESS MAIL - Label No.** DEK938097140US

Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

Sir:

Enclosed herewith for filing is a patent application of YANNICK BATARD, FRANCIS DURST, MICHEL SCHALK and DANIELE WERCK-REICHHART entitled RECODING OF DNA SEQUENCES PERMITTING EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST

which includes:

| | |
|---|---|
| [X] Specification | 42 Total Pages |
| [X] Claims | 6 Total Pages |
| [X] Abstract | 1 Total Pages |
| [] Drawing(s) | ___ Total Sheets |
| _ formal | |
| _ informal | |

[X] Combined Declaration and Power of Attorney 3 Total Pages
[] Newly executed (original or copy)
[X] Copy from a prior application
(for continuation/divisional only - **must be filed to avoid surcharge for late filing)**

If a continuing application, check appropriate box:
[X] Continuation [] Divisional [] Continuation-In-Part (CIP)
of prior application No. 09/158,767

[X] Amend the specification by inserting, before the first line, the following sentence:

"This is a [X] continuation [] divisional [] continuation-in-part of copending application Serial No. 09/158,767 filed September 23, 1998."

Appln. Trans.
PATENT

Attorney Docket No. A32000-A-072667.0172

[X] An Assignment of the invention to RHONE-POULENC AGRO .
 [] is attached. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
 [] will follow.
 [X] has been filed in the prior application

[] Small Entity Statement(s) **ENCLOSED**.
[] Small Entity Statement filed in prior application. Status still proper and desired.

[X] Information Disclosure Statement (IDS) PTO-1449
 [X] Copies of IDS Citations.

[X] Preliminary Amendment

[X] Return Receipt Postcard

[X] Other Letter Under 37 C.F.R. 1.821(e)

[] Cancel in this application original claims _ of the prior application before calculating the filing fee.

The filing fee has been calculated as shown below:

| FOR | (Col. 1) No.Filed | | | (Col. 2) No. Extra | Small Entity Rate | Small Entity Fee | OR | Other Than A Small Entity Rate | Other Than A Small Entity Fee |
|---|---|---|---|---|---|---|---|---|---|
| Basic Fee | | | | | | | | | $710.00 |
| Total Claims | 28 | -20 | = | 8 | x 9 = | $0.00 | | x 18 = | $144.00 |
| Ind. Claims | 2 | -3 | = | 0 | x 40 = | $0.00 | | x 80 = | $0.00 |
| Multiple Dependent Claim | | | | | + 135 = | | | + 270 = | |
| | | | | | Total | $0.00 | | | $854.00 |

* If the difference in Col. 1 is less than zero, enter "0" in Col. 2.
 Fee Payment Being Made:
 [X] Enclosed

| | | |
|---|---|---|
| [X] | Basic filing fee | $854.00 |
| [] | Recording Assignment [$40.00; 37 CFR 1.21(h)] | $0.00 |
| | Total Fees Enclosed | $854.00 |

[X] A check in the amount of $854.00 to cover filing fee is enclosed.

Appln. Trans.
PATENT

Attorney Docket No. A32000-A-072667.0172

Priority

[X] Priority of application Country FRANCE, Appln. No. 9712094 filed September 24, 1997 is claimed under 35 U.S.C. 119.

[X] Certified Copy of Priority Document(s) Country FRANCE, Appln No. 9712094, filed September 24, 1997.

[] is/are attached [] will follow [X] has been filed in the parent application S/N 09/158,767.

[X] The Commissioner is hereby authorized to charge payment of any additional filing fees required under 37 CFR 1.16, 1.17, and 1.21(h) associated with this communication or credit any overpayment to Deposit Account No. 02-4377. Two copies of this sheet are enclosed.

BAKER BOTTS L.L.P.

By Janet M. MacLeod
Janet M. MacLeod
PTO Registration No. 35,263

Enclosures

FILE NO. A32000-A-072667.0172

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

| | | | | |
|---|---|---|---|---|
| Applicant | : | Yannick Batard et al. | | |
| Serial No. | : | NOT YET ASSIGNED | Examiner: | |
| Filed | : | HEREWITH | Group Art Unit: | |
| For | : | RECODING OF DNA SEQUENCES PERMITTING EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST | | |

**<u>PRELIMINARY AMENDMENT</u>**

Assistant Commissioner for Patents

Washington, D.C. 20231

Sir:

Please amend the above-identified application as follows:

<u>IN THE SPECIFICATION</u>:

<u>Page 12, lines 15-16</u>, delete "(sequence identifier No. 1)" and substitute therefor --of SEQ ID NO: 1 (which encodes the amino acid sequence of SEQ ID NO: 15)--.

Page 14, line 11, after "No. 7" insert --(which encodes the amino acid sequence of SEQ. ID NO: 16)--.

Page 14, line 11, after "No. 8" insert --(which encodes the amino acid sequence of SEQ. ID NO: 17)--.

Page 14, line 11, after "No. 9" insert --(which encodes the amino acid sequence of SEQ ID NO: 18)--.

Page 18, line 2, after "No. 10" insert --, which encodes the amino acid sequence of SEQ ID NO: 19--.

Page 18, line 14, after "No. 14" insert --, which encodes the amino acid sequence of SEQ ID NO: 20--.

Please delete pages 20-42 and renumber Pages 43-48 as pages 20-25.

After page 48, please insert the attached substitute sequence listing.

IN THE CLAIMS:

Claim 5, lines 1-2, delete “one of Claims 1 to 4" and substitute therefor --claim 1--.

Claim 7, lines 1-2, delete “one of claims 1 to7" and substitute therefor --claim 1--.

Claim 11, lines 1-2, delete “one of claims 9 or 10" and substitute therefor

--claim 9--..

Claims 12, lines 1-2, delete “one of claims 1 to 11" and substitute therefor

--claim 1--.

Clam 13, lines 1-2, delete “one of claims 1 to 12" and substitute therefor

--claim 1--.

Claim 15, lines 1-2, delete “one of claims 1 to 14" and substitute therefor

--claim 1--.

Claim 18, lines 1-2, delete “one of claims 1 to 17" and substitute therefor

--claim 1--.

Claim 22, line 2, delete “one of claims 1 to 21" and substitute therefor

--claim 1--.

Claim 27, line 5, delete “according to claim 23".

Claim 27, line 6, delete “one of  claims 1 to 21" and substitute therefor

--claim 1--.

Claim 28, line 6, delete “according to claim 23".

Claim 28, lines 7-8, delete “one of claims 1 to 21" and substitute therefor

--claim 1--.

**REMARKS**

The foregoing amendments are necessary to conform the specification to the Sequence Listing and to remove multiple dependencies. No new matter has been introduced by the foregoing amendments.

Respectfully submitted,

Janet M. MacLeod

Louis S. Sorell
Patent Office Reg. No. 32,439

Attorney for Applicants
(212) 408-2620

Janet M. MacLeod
Patent Office Reg. No. 35,263

Attorney for Applicants
(212) 408-2597

# The recoding of DNA sequences to enable them to be expressed in yeasts, and the transformed yeasts obtained

The present invention relates to the recoding of DNA sequences which encode proteins which contain regions having a high content of codons which are poorly translated by yeasts, in particular which encode proteins of plant origin, such as the P450 cytochromes of plant origin, and to their expression in yeasts.

It is known that certain sequences encoding proteins of interest, in particular proteins of plant origin, are not readily translated in yeasts. This applies, in particular, to proteins which possess regions having a high content of codons which are poorly suited to yeasts, in particular leucine codons, such as some P450 cytochromes of plant origin. Some systems which have been developed for improving the expression of P450 cytochromes of animal or plant origin in yeasts, such as those described by Pompon et al. (*Methods Enzymol.*, 272, 1996, 51-64; WO 97/10344), have turned out to be unsuitable for large numbers of P450 cytochromes which encompass regions having a high content of codons which are poorly suited to yeasts.

The P450 cytochromes constitute a superfamily of membrane enzymes of the monooxygenase type which are able to oxidize a large family of generally hydrophobic substrates. The reactions are most frequently characterized by the oxidation of C-H or C=C bonds, and

of heteroatoms, and, more rarely, by the reduction of nitro groups or by dehalogenation. More specifically, these enzymes are involved in the metabolism of xenobiotic substances and drugs and in the biosynthesis of secondary metabolites in plants, some of which have organoleptic or pharmacodynamic properties.

As a consequence, the P450 cytochromes are used, in particular, in:

- the *in vitro* diagnosis of the formation of toxic or mutagenic metabolites (molecules of natural origin, pollutants, drugs, pesticides, etc.), making it possible, in particular, to develop novel active molecules (pharmaceutical, agrochemistry),
- the identification and destruction of molecules which are toxic for, or pollute, the environment,
- the enzymic synthesis of novel molecules.

The search for heterologous expression of P450 cytochromes by host cells, more specifically yeasts, is therefore important for obtaining controlled production of this enzyme in large quantity, either for isolating it and using it in the above-listed processes, or for using the transformed cells directly for the said processes without previously isolating the enzyme.

The present invention provides a solution to the abovementioned problem, enabling proteins which contain regions having a high content of codons which

are poorly suited to yeasts, in particular P450 cytochromes of plant origin, to be expressed in yeasts.

The present invention therefore relates to a DNA sequence, in particular a cDNA sequence, which encodes a protein of interest which contains regions having a high content of codons which are poorly suited to yeasts, characterized in that a sufficient number of codons which are poorly suited to yeasts is replaced with corresponding codons which are well-suited to yeasts in the said regions having a high content of codons which are poorly suited to yeasts.

Within the meaning of the present invention, "codons which are poorly suited to yeasts" are understood as being codons whose frequency of use by yeasts is less than or equal to approximately 13 per 1000, preferably less than or equal to approximately 12 per 1000, more preferably less than or equal to approximately 10 per 1000. The frequency at which codons are used by yeasts, more specifically by *S. cerevisiae*, is described, in particular, in "Codon usage data base from Yasukazu Nakamura" (http://www.dna.affrc.go.jp/~nakamura/codon.html). This applies, in particular, to codons CTC, CTG and CTT, which encode leucine, to codons CGG, CGC, CGA, CGT and AGG, which encode arginine, to codons GCG and GCC, which encode alanine, to codons GGG, GGC and GGA, which encode glycine, and to codons CCG and CCC, which encode proline. The codons which are poorly suited to yeasts

in accordance with the invention are, more specifically, codons CTC and CTG, which encode leucine, CGG, CGC, CGA, CGT and AGG, which encode arginine, codons GCG and GCC, which encode alanine, GGG and GGC, which encode glycine, and codons CCG and CCC, which encode proline.

Within the meaning of the present invention, "corresponding codons which are well-suited to yeasts" are understood as being the codons which correspond to the codons which are poorly suited to yeasts and which encode the same amino acids, and whose frequency of use by yeasts is greater than 15 per 1000, preferably greater than or equal to 18 per 1000, more preferably greater than or equal to 20 per 1000. This applies, in particular, to codons TTG and TTA, preferably TTG, which encode leucine, to codon AGA, which encodes arginine, to codons GCT and GCA, preferably GCT, which encode alanine, to codon GGT, which encodes glycine, and to codon CCA, which encodes proline.

Within the meaning of the present invention, "region having a high content of codons which are poorly suited to yeasts" is understood as being any region of the DNA sequence which contains at least 2 poorly suited codons among 10 consecutive codons, with it being possible for the two codons to be adjacent or separated by up to 8 other codons. According to one preferred embodiment of the invention, the regions having a high content of poorly suited codons contain

2, 3, 4, 5 or 6 poorly suited codons per 10 consecutive codons, or contain at least 2 or 3 adjacent poorly suited codons.

Within the meaning of the present invention, "sufficient number of codons" is understood as being the number of codons which it is necessary and sufficient to replace in order to observe a substantial improvement in their expression in yeasts. Advantageously, at least 50% of the codons which are poorly suited to yeasts in the high-content region under consideration are replaced with well-suited codons. Preferably, at least 75% of the poorly suited codons of the said region are replaced, with 100% of the poorly suited codons more preferably being replaced.

Within the meaning of the present invention, "substantial improvement" is understood as being either a detectable expression when no expression of the reference sequence is observed, or an increase in expression as compared with the level at which the reference sequence is expressed.

Within the meaning of the present invention, "reference sequence" designates any sequence which encodes a protein of interest and which is modified in accordance with the invention in order to promote its expression in yeasts.

The present invention is particularly well suited to DNA sequences, in particular cDNA sequences,

which encode proteins of interest which contain regions having a high content of leucine and in which a sufficient number of CTC codons encoding leucine in the said region having a high content of leucine is replaced with TTG and/or TTA codons, or in which a sufficient number of CTC and CTG codons encoding leucine in the said region having a high content of leucine is replaced with TTG and/or TTA codons, preferably with a TTG codon.

Within the meaning of the present invention, "region having a high content of leucine" is understood as being a region which contains at least 2 leucines among 10 consecutive amino acids in the protein of interest, with it being possible for the two leucines to be adjacent or separated by up to 8 other amino acids. According to one preferred embodiment of the invention, the regions having a high content of leucine contain 2, 3, 4, 5 or 6 leucines per 10 consecutive amino acids, or contain at least 2 or 3 adjacent leucines.

According to a preferred embodiment of the invention, at least 50% of the CTC or CTC and CTG codons of the region having a high content of leucine are replaced with TTG or TTA codons, with at least 75% of the CTC or CTC and CTG codons of the said region preferably being replaced, and 100% of the CTC or CTC and CTG codons more preferably being replaced.

Advantageously, the present invention is

particularly suitable for DNA sequences whose general content of poorly suited codons is at least 20%, more preferably at least 30%, as compared with the total number of codons in the reference sequence.

Advantageously, when the reference sequence contains at least one 5′ region having a high content of poorly suited codons, the recoding of this 5′ region alone makes it possible to obtain a substantial improvement in the expression of the protein of interest in yeasts. The length of the 5′ region to be recoded in accordance with the invention will vary depending on the length of the region having a high content of poorly suited codons. This length will advantageously be at least four codons, in particular when this region contains at least two adjacent poor codons, up to approximately 40 codons or more.

However, it is not necessary, according to the invention, to recode all the reference sequence, but only the regions having a high content of poor codons, in particular the 5′ region on its own, in order to obtain a substantial improvement in the expression of the protein of interest in yeasts.

Advantageously, the DNA sequence encoding a protein of interest is an isolated DNA sequence of natural origin, in particular of plant origin. The invention is particularly advantageous for sequences which originate from monocotyledonous or dicotyledonous plants, preferably monocotyledonous plants, in

particular of the graminae family, such as wheat, barley, oats, rice, maize, sorghum, cane sugar, etc.

According to a preferred embodiment of the invention, the DNA sequence encodes an enzyme, in particular a cytochrome P450, which is preferably of plant origin. These P450 cytochromes exhibit a high content of poorly suited codons, in particular encoding leucine, in their N-terminal region; it is in the 5′-terminal coding region that the poorly suited codons are replaced.

The present invention also relates to a chimeric gene which comprises a DNA sequence which has been modified as above and heterologous 5′ and 3′ regulatory elements which are able to function in a yeast, that is to say which are able to control the expression of the protein of interest in the yeast. Such regulatory elements are well known to the skilled person and are described, in particular, by Rozman et al. (Genomics, 38, 1996, 371-381) and by Nacken et al. (Gene, 175, 1996, 253-260, *Probing the limits of expression levels by varying promoter strength and plasmid copy number in Saccharomyces cerevisiae*).

The present invention also relates to a vector for transforming yeasts which contains at least one chimeric gene as described above. It also relates to a process for transforming yeasts with the said vector and to the transformed yeasts which are obtained. It finally relates to a process for producing

a heterologous protein of interest in a transformed yeast, with the sequence which encodes the said protein of interest being such as defined above.

The process for producing a heterologous protein of interest in a transformed yeast comprises the steps of:

a) transforming a yeast with a vector which is able to replicate in yeasts and which contains a modified DNA sequence as defined above and heterologous 5′ and 3′ regulatory elements which are able to function in a yeast,

b) culturing the transformed yeast, and

c) extracting the protein of interest from the yeast culture.

When the protein of interest is an enzyme which is suitable for transforming a substrate, such as a cytochrome P450, the enzyme which has been extracted from the yeast culture is then used for catalysing the transformation of the said substrate.

However, the catalysis can be carried out, without requiring the extraction of the yeast, by culturing the transformed yeast in the presence of the said substrate.

The present invention also relates, therefore, to a process for transforming a substrate by enzymic catalysis using an enzyme which is expressed in a yeast, which process comprises the steps of

a) culturing the yeast which has been

transformed in accordance with the invention in the presence of the substrate to be transformed, then

b) recovering the transformed substrate from the yeast culture.

When the yeast has been transformed for expressing a cytochrome P450, the reaction which is catalysed by the enzyme is an oxidation reaction, more specifically a reaction in which C-H or C=C bonds are oxidized.

The techniques for transforming and culturing yeasts are known to the skilled person, and are described, for example, in *Methods in Enzymology* (Vol. 194, 1991).

Yeasts which are of use in accordance with the invention are selected, in particular, from the genera *Saccharomyces, Kluyveromyces, Hansenula, Pichia* and *Yarrowia*. Advantageously, the yeast belongs to the *Saccaromyces* genus, and is in particular *S. cerevisiae*.

Other characteristics of the invention will become apparent in the light of the examples which follow.

**Example 1: Production of a wheat cDNA gene library, and identification of the CYP73A17 sequence**

The wheat cytochrome P450 CYP73A17 sequence was obtained by screening a young wheat plantlet (shoots and roots without the caryopses) cDNA library which was constructed in the vector λ-ZapII (Stratagene) in accordance with the supplier's

instructions.

**1. Production of the cDNA library**

*Triticum aestivum* (L. cv. Darius) seeds which had been coated with cloquintocet-mexyl (0.1% per dry weight of seed) are cultured in plastic boxes on two layers of damp gauze until shoots having a size of 3 to 5 mm are obtained. The water in the boxes is then replaced with a solution of 4 mM sodium phenobarbital and the wheat is cultured until the shoots are approximately 1 cm in size.

The cDNA library is constructed in the λ-ZapII (Stratagene) vector, in accordance with the supplier's protocol and instructions, using 5 μg of poly(A)$^+$ RNA (Lesot, A., Benveniste, I., Hasenfratz, M.P., Durst, F. (1990) Induction of NADPH cytochrome P450(c) reductase in wounded tissues from *Helianthus tuberosus* tubers. Plant Cell Physiol., 31, 1177-1182) which were isolated from the treated roots and shoots.

**2. Screening the cDNA library**

$5\times10^5$ lysis plaques from the previously obtained λ-ZapII library are screened using a probe which corresponds to the complete coding sequence of *Helianthus tuberosus* CYP73A1, and which has been labelled by random priming with [$\alpha$-$^{32}$P]dCTP. The filters are prehybridized and hybridized at low stringency at 55°C in accordance with the standard protocols. The membranes are washed twice for 10 minutes with 2 × SSC, 0.1% SDS, and once for 10 minutes with 0.2 × SSC, 0.1%

SDS at ambient temperature, then twice for 30 minutes with 0.2 × SSC, 0.1% SDS at 45°C. The inserts of the positive lysis plaques are analysed by PCR (polymerization chain reaction) and hybridization in order to determine their size. The clones containing inserts which hybridize with CYP73A1 under the above-described conditions and which are greater than 1.5 kbp in size are rescreened before excision of the pBluescript plasmid in accordance with the supplier's (Stratagene) protocol and sequencing using the Ready Reaction Dye Deoxy Terminator Cycle prism technique developed by Applied Biosystems Inc. A full length clone is then identified by alignment with CYP73A1.

The wheat cytochrome P450 CYP73A17 which is encoded by the isolated sequence (sequence identifier No. 1) exhibits 76.2% identity with the *Helianthus tuberosus* CYP73A1.

**Example 2: Alterations to the sequence encoding the wheat cytochrome P450 CYP73A17**

Contrary to the situation with regard to *Helianthus tuberosus* CYP73A1, which can be expressed in yeasts (Urban et al., 1994), repeated attempts to express wheat CYP73A17 in yeasts using the same customary techniques proved to be fruitless when the nucleotide sequence was not altered at the time it was inserted into the expression vector (verification by sequencing). No protein is detected by spectrophotometry or by immunoblotting, just as no

enzymic activity is detectable in the microsomes of transformed and induced yeast.

**1. Alteration of the coding sequence**

The sequence encoding wheat CYP73A17 (SEQ. ID No. 1) was therefore altered, in three different ways, by PCR-induced mutagenesis, as follows:

The *BamHI* and *EcoRI* restriction sites were respectively introduced by PCR just upstream of the ATG codon and just downstream of the stop codon of the CYP73A17 coding sequence (source, origin) using the sense and reverse primers described below, with the restriction sites being *BamHI* in the case of the sense primers Rec1 (SEQ ID No. 3), Rec2 (SEQ ID No. 4) and Rec3 (SEQ ID No. 5), and *EcoRI* in the case of the reverse primer (SEQ ID No. 6).

A primer, represented by SEQ ID No. 2, was also employed for enabling yeasts to be transformed with the unmodified (native) sequence encoding wheat CYP73A17.

The five primers described above were obtained from Eurogentech, and were synthesized and purified in accordance with customary methods.

For each alteration using the four different sense primers, the mode of operation is as follows:

The reaction mixture (20 mM Tris-HCl, pH 8.75, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X100, 0.1 mg/ml BSA, 5% (v/v) DMSO, 300 $\mu$M dNTP, 20 pmoles of each primer, 150 ng of template, total

volume 50 $\mu$l) is preheated at 94°C for 2 minutes before adding 5 units of Pfu DNA polymerase (Stratagene). After 2 minutes at 94°C, 30 amplification cycles are carried out as follows: 1 minute of denaturation at 94°C, 2 minutes of hybridization at 55°C, 2 minutes of extension at 72°C. The reaction is completed by 10 minutes of extension at 72°C.

For each primer, a sequence is obtained which is derived from sequence ID No. 1, and which is represented, in the case of the altered coding sequences, by the sequences ID No. 7, No. 8 and No. 9. The 5′ ends of the sequences obtained using the four abovementioned sense primers are depicted below, with the *BamHI* restriction site being shown in italics:

| | | | |
|---|---|---|---|
| native: | ATATAT*GGATCC* | ATG GAC GTC CTC CTC CTG GAG AAG GCC | CTC CTG GGC CTC TTC GCC GCG GCG GTG CTG GCC ATC GCC GTC GCC |
| Rec 1 | ATATAT*GGATCC* | ATG GAT GTT TTG TTG TTG GAG AAG GCC | CTC CTG GGC CTC TTC GCC GCG GCG GTG CTG GCC ATC GCC GTC GCC |
| Rec 2 | ATATAT*GGATCC* | ATG GAT GTT TTG TTG TTG GAA AAA GCT | TTG TTG GGT TTG TTC GCC GCG GCG GTG CTG GCC ATC GCC GTC GCC |
| Rec 3 | ATATAT*GGATCC* | ATG GAT GTT TTG TTG TTG GAA AAA GCT | TTG TTG GGT TTG TTT GCT GCT GCT GTT TTG GCT ATT GCT GTT GCT |
| Protein: | | met asp val leu leu leu glu lys ala | leu leu gly leu phe ala ala ala val leu ala ile ala val ala |

**AAG CTC ACC GGC AAG CGC TTC CGC CTC CCC CCT GGC CCC TCC GGC**
**AAG CTC ACC GGC AAG CGC TTC CGC CTC CCC CCT GGC CCC TCC GGC**
**AAG CTC ACC GGC AAG CGC TTC CGC CTC CCC CCT GGC CCC TCC GGC**
**AAA TTG ACT GGT AAA AGA TTT AGA TTG CCA CCA GGT CCA TCC GGC**
**lys leu thr gly lys arg phe arg leu pro pro gly pro ser gly**

**GCC CCC ATC GTC .....**
**GCC CCC ATC GTC .....**
**GCC CCC ATC GTC .....**
**GCC CCC ATC GTC .....**
**ala pro ile val .....**

**2. Transforming the yeasts**

After having been digested with the restriction enzymes *BamHI* and *EcoRI*, the four above-described altered coding sequences are integrated into the vector pYeDP60, which is described by Pompon et al. (*Methods Enzymol*, 272, 1996, 51-64; WO 97/10344), the content of which is hereby incorporated by reference with regard to the plasmid, the method of insertion into the plasmid, and the method of transforming and growing the yeasts, in particular using the *Saccharomyces cerevisiae* yeast strains W(R), WAT21 and WAT11. The method for transforming and growing yeasts is also described by Pompon et al. and by Urban et al.

(*Eur. J. Biochem,* 222, 1994, page 844, 2nd column, "Yeast transformation and cell culture").

4 transformed yeast strains, designated: W73A17(native), W73A17(Rec1), W73A17(Rec2) and W73A17(Rec3), are obtained.

**Example 3: Expression of CYP73A17 in the altered yeasts**

The previously obtained transformed yeasts are cultured, in accordance with the method described by Urban et al. (*Eur. J. Biochem.*, 222, 1994, page 844, 2nd column, "Yeast transformation and cell culture"), in 50 ml of SGI medium at 30°C for 72 h. The cells are recovered by centrifuging at 8000 g for 10 minutes, washed with 25 ml of YPI medium, recentrifuged, and then resuspended in 250 ml of YPI medium. The cells are induced with galactose for 14-16 h, while being shaken at 160 rpm, until the cell density reaches $10^8$ cells per ml. The microsomes are then prepared using the method described by Pierrel et al. (*Eur. J. Biochem.*, 224, 1994, 835-844).

The expression of CYP73A17 achieved in the case of the four strains is quantified by differential spectrophotometry using the method described by Omura and Sato (*J. Biol. Chem., 177, 678-693*). It is proportional to the number of poorly suited codons which have been altered.

The microsomal enzymic activity is measured using the method described by Durst F., Benveniste I., Schalk M. and Werck-Reichhart D. (1996) Cinnamic acid

hydroxylase activity in plant microsomes. Methods Enzymol. 272, 259-268. The results obtained after transforming WAT21 are recorded in the Table below. The activity is expressed as cinnamate 4-hydroxylase activity. The percentage additional activity (rounded values) illustrates the extent of the leap in activity which is observed after the poorly suited codons have been altered.

| Strain | Activity pmol/min/μg of protein | % additional activity |
|---|---|---|
| W73A17 native | 0.64 | - |
| W73A17 Rec1 | 2.84 | +340 |
| W73A17 Rec2 | 4.92 | +670 |
| W73A17 Rec3 | 8.90 | +1300 |

These results relating to the increase in enzymic activity confirm those relating to the increase in the expression of the protein in the yeasts. They demonstrate that alteration of the 5′ end alone, even when limited (Rec1), is sufficient to obtain a very substantial improvement in the production of the enzyme by the yeast and in its enzymic activity.

**Example 4: Expression of wheat CYP86A5 in the altered yeasts**

The sequence encoding wheat cytochrome P450

CYP86A5, which is depicted by sequence identifier No. 10 (SEQ ID No. 10), was isolated from the wheat cDNA library described in Example 1 using the same method of operation as described for the CYP73A17 sequence and employing the complete coding sequence of *Arabidopsis thaliana* CYP86A1 as the probe. This wheat CYP86A5 sequence was altered, in accordance with the mode of operation of Example 2, using the two oligonucleotides depicted by the sequences ID No. 12 and 13 (SEQ ID No. 12 and SEQ ID No. 13) as sense and reverse primers, respectively, in order to obtain the coding sequence which is altered in accordance with the invention and which is depicted by sequence identifier No. 14 (SEQ ID No. 14).

A primer depicted by SEQ ID No. 11 was also used to enable yeasts to be transformed with the sequence encoding unmodified (native) wheat CYP86A5.

The yeasts are transformed with this new coding sequence and the expression is quantified by differential spectrophotometry in accordance with the mode of operation described in Example 2. While the natural sequence of wheat CYP86A5 is not expressed in a detectable manner, there is substantial expression in the transformed yeasts of the sequence which has been modified in accordance with the invention.

The above-described examples demonstrate unambiguously that the expression in yeasts of DNA sequences which possess a 5′ region having a high

content of codons which are poorly suited to yeasts is substantially improved when this region alone is simply recoded in accordance with the invention, ever partially, with corresponding codons which are well-suited to yeasts.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 14

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2261 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 49..1551

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
CGCAGCACGG CAACACATAC ACAGGAGCCA CACACCGCAC CTACCCCG ATG GAC GTC      57
                                                     Met Asp Val
                                                      1

CTC CTC CTG GAG AAG GCC CTC CTG GGC CTC TTC GCC GCG GCG GTG CTG      105
Leu Leu Leu Glu Lys Ala Leu Leu Gly Leu Phe Ala Ala Ala Val Leu
     5                  10                  15

GCC ATC GCC GTC GCC AAG CTC ACC GGC AAG CGC TTC CGC CTC CCC CCT      153
Ala Ile Ala Val Ala Lys Leu Thr Gly Lys Arg Phe Arg Leu Pro Pro
 20                  25                  30                  35

GGC CCC TCC GGC GCC CCC ATC GTC GGC AAC TGG CTG CAG GTC GGC GAC      201
Gly Pro Ser Gly Ala Pro Ile Val Gly Asn Trp Leu Gln Val Gly Asp
                 40                  45                  50

GAC CTC AAC CAC CGC AAC CTG ATG GGC CTG GCC AAG CGG TTC GGC GAG      249
Asp Leu Asn His Arg Asn Leu Met Gly Leu Ala Lys Arg Phe Gly Glu
             55                  60                  65

GTG TTC CTC CTC CGC ATG GGC GTC CGC AAC CTG GTG GTC GTC TCC AGC      297
Val Phe Leu Leu Arg Met Gly Val Arg Asn Leu Val Val Val Ser Ser
         70                  75                  80

CCC GAG CTC GCC AAG GAG GTC CTC CAC ACC CAG GGC GTC GAG TTC GGC      345
Pro Glu Leu Ala Lys Glu Val Leu His Thr Gln Gly Val Glu Phe Gly
     85                  90                  95

TCC CGC ACC CGC AAC GTC GTC TTC GAC ATC TTC ACC GGC AAG GGA CAG      393
Ser Arg Thr Arg Asn Val Val Phe Asp Ile Phe Thr Gly Lys Gly Gln
100                 105                 110                 115

GAC ATG GTG TTC ACG GTG TAC GGC GAC CAC TGG CGC AAG ATG CGG CGG      441
Asp Met Val Phe Thr Val Tyr Gly Asp His Trp Arg Lys Met Arg Arg
                120                 125                 130

ATC ATG ACG GTG CCC TTC TTC ACC AAC AAG GTG GTG GCG CAG AAC CGC      489
Ile Met Thr Val Pro Phe Phe Thr Asn Lys Val Val Ala Gln Asn Arg
            135                 140                 145

GTG GGG TGG GAG GAG GAG GCC CGG CTG GTG GTG GAG GAC CTC AAG GCC      537
Val Gly Trp Glu Glu Glu Ala Arg Leu Val Val Glu Asp Leu Lys Ala
        150                 155                 160
```

GAC CCG GCG GCG GCG ACG GCG GGC GTG GTG GTC CGC CGC AGG CTG CAG 585
Asp Pro Ala Ala Ala Thr Ala Gly Val Val Val Arg Arg Arg Leu Gln
165 170 175

CTC ATG ATG TAC AAC GAC ATG TTC CGC ATC ATG TTC GAC CGC CGG TTC 633
Leu Met Met Tyr Asn Asp Met Phe Arg Ile Met Phe Asp Arg Arg Phe
180 185 190 195

GAG AGC GTG GCC GAC CCG CTC TTC AAC CAG CTC AAG GCG CTC AAC GCC 681
Glu Ser Val Ala Asp Pro Leu Phe Asn Gln Leu Lys Ala Leu Asn Ala
200 205 210

GAG CGC AGC ATC CTC TCC CAG AGC TTC GAC TAC AAC TAC GGC GAC TTC 729
Glu Arg Ser Ile Leu Ser Gln Ser Phe Asp Tyr Asn Tyr Gly Asp Phe
215 220 225

ATC CCC GTC CTC CGC CCC TTC CTC CGC CGC TAC CTC AAC CGC TGC ACC 777
Ile Pro Val Leu Arg Pro Phe Leu Arg Arg Tyr Leu Asn Arg Cys Thr
230 235 240

AAC CTC AAG ACC AAG CGG ATG AAG GTG TTC GAG GAC CAC TTC GTC CAG 825
Asn Leu Lys Thr Lys Arg Met Lys Val Phe Glu Asp His Phe Val Gln
245 250 255

CAG CGC AAG GAG GCG TTG GAG AAG ACG GGT GAG ATC AGG TGC GCC ATG 873
Gln Arg Lys Glu Ala Leu Glu Lys Thr Gly Glu Ile Arg Cys Ala Met
260 265 270 275

GAC CAC ATC CTG GAA GCC GAA AGG AAG GGC GAG ATC AAC CAC GAC AAC 921
Asp His Ile Leu Glu Ala Glu Arg Lys Gly Glu Ile Asn His Asp Asn
280 285 290

GTC CTC TAC ATC GTC GAG AAC ATC AAC GTC GCA GCC ATC GAG ACG ACG 969
Val Leu Tyr Ile Val Glu Asn Ile Asn Val Ala Ala Ile Glu Thr Thr
295 300 305

CTG TGG TCG ATC GAG TGG GGC CTC GCG GAG CTG GTG AAC CAC CCG GAG 1017
Leu Trp Ser Ile Glu Trp Gly Leu Ala Glu Leu Val Asn His Pro Glu
310 315 320

ATC CAG CAG AAG CTG CGC GAG GAG ATC GTC GCC GTT CTG GGC GCC GGC 1065
Ile Gln Gln Lys Leu Arg Glu Glu Ile Val Ala Val Leu Gly Ala Gly
325 330 335

GTG GCG GTG ACG GAG CCG GAC CTG GAG CGC CTC CCC TAC CTG CAG TCC 1113
Val Ala Val Thr Glu Pro Asp Leu Glu Arg Leu Pro Tyr Leu Gln Ser
340 345 350 355

GTG GTG AAG GAG ACG CTC CGC CTC CGC ATG GCA ATC CCG CTC CTG GTG 1161
Val Val Lys Glu Thr Leu Arg Leu Arg Met Ala Ile Pro Leu Leu Val
360 365 370

CCG CAC ATG AAC CTC AGC GAC GCC AAG CTC GCC GGC TAC GAC ATC CCC 1209
Pro His Met Asn Leu Ser Asp Ala Lys Leu Ala Gly Tyr Asp Ile Pro
375 380 385

GCC GAG TCC AAG ATC CTC GTC AAC GCC TGG TTC CTC GCC AAC GAC CCC 1257
Ala Glu Ser Lys Ile Leu Val Asn Ala Trp Phe Leu Ala Asn Asp Pro
390 395 400

AAG CGG TGG GTG CGC GCC GAT GAG TTC AGG CCG GAG AGG TTC CTC GAG 1305
Lys Arg Trp Val Arg Ala Asp Glu Phe Arg Pro Glu Arg Phe Leu Glu
405 410 415

GAG GAG AAG GCC GTC GAG GCC CAC GGC AAC GAT TTC CGG TTC GTG CCC 1353
Glu Glu Lys Ala Val Glu Ala His Gly Asn Asp Phe Arg Phe Val Pro
420 425 430 435

```
TTC GGC GTC GGC CGC CGG AGC TGC CCC GGG ATC ATC CTC GCG CTG CCC      1401
Phe Gly Val Gly Arg Arg Ser Cys Pro Gly Ile Ile Leu Ala Leu Pro
                440                 445                 450

ATC ATC GGC ATC ACG CTC GGA CGC CTG GTG CAG AAC TTC CAG CTG CTG      1449
Ile Ile Gly Ile Thr Leu Gly Arg Leu Val Gln Asn Phe Gln Leu Leu
            455                 460                 465

CCG CCG CCG GGG CAG GAC AAG ATC GAC ACC ACC GAG AAG CCC GGG CAG      1497
Pro Pro Pro Gly Gln Asp Lys Ile Asp Thr Thr Glu Lys Pro Gly Gln
        470                 475                 480

TTT ACC AAC CAG ATC CTC AAG CAC GCC ACC ATT GTC TGC AAG CCA CTC      1545
Phe Thr Asn Gln Ile Leu Lys His Ala Thr Ile Val Cys Lys Pro Leu
    485                 490                 495

GAG GCT TAACTGAATT GAGGTTTCGG TCATGGGCGC CCGCTGACGC GGGGAGATGG       1601
Glu Ala
500

ATCTATGCAT GTGACTGTGT ATTTTGCCTT CTTTCTTTTT GGTGTTGTTT TTTGCAGTAG    1661

TAAGTTTAAT TTTTCTTTGG TGTTGGCCTA TTTGTCTTCA TGTGAGGCGT CGTGTTGTAA    1721

ATTTCCATAT AGTTGGCAAT GTGATGTAAA ACTTGGCTCC AAAAAAAAAA AAAAAAAACT    1781

CGAGACTCTT CTCTCTCTCT CTCTCTCTCC AGCCTCGGGT CTCTGCTGGC AAGGGAACTT    1841

GCATTACCCT GTGTACGACG GCGCCATGTT CGTCCCTGAA GCACCCTCCC TGCAGAGCTC    1901

CCAGGACAAC TTCGCTGCAT CTGCTGGTTT CAAGCGTCGA AGGAGAGAGT TTTGAATACC    1961

CGAAAGAATA TAGCGTTGGA CATATCTGTC AAACAGGGGA TCTTGCTGTG GGTCTCTTGG    2021

TGGGCCAAAT CGCATAGACA ATCATTCAAA TGGATGGGTT CTTCGCTGGT CGGTCAAAAA    2081

GTATATGTTG TAATTGTACG CCTTTTTTGG GTCTTGTTGC CAAAGATCAT GGTTATTGAG    2141

TTGTGAGCTC TGAGATAACA GGTTTGTGTA TAGTGAAATA AAGAGGAGCG TCGTCAACAC    2201

CATGTACTAT ATAGGCTTTG AAATTCCATT AAGATGCATC AGAAATCAAT GTTGGATTTG    2261
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

**ATATATGGAT CCATGGACGT CCTCCTCCTG GAGAAGGC** 38

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

**ATATATGGAT CCATGGATGT TTTGTTGTTG GAGAAGGCCC TCCTGGGCCT CTTCGC** 56

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 71 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

ATATATGGAT CCATGGATGT TTTGTTGTTG GAAAAAGCTT TGTTGGGTTT GTTCGCCGCG 60
GCGGTGCTGG C 71

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 143 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

ATATATGGAT CCATGGATGT TTTGTTGTTG GAAAAAGCTT TGTTGGGTTT GTTTGCTGCT 60
GCTGTTTTGG CTATTGCTGT TGCTAAATTG ACTGGTAAAA GATTTAGATT GCCACCAGGT 120
CCATCCGGCG CCCCCATCGT CGG 143

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TATATAGAAT TCCAGTTAAG CCTCGAGTGG CTTGCAGAC 39

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1506 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1503

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
ATG GAT GTT TTG TTG TTG GAG AAG GCC CTC CTG GGC CTC TTC GCC GCG      48
Met Asp Val Leu Leu Leu Glu Lys Ala Leu Leu Gly Leu Phe Ala Ala
1               5                   10                  15

GCG GTG CTG GCC ATC GCC GTC GCC AAG CTC ACC GGC AAG CGC TTC CGC      96
Ala Val Leu Ala Ile Ala Val Ala Lys Leu Thr Gly Lys Arg Phe Arg
            20                  25                  30

CTC CCC CCT GGC CCC TCC GGC GCC CCC ATC GTC GGC AAC TGG CTG CAG     144
Leu Pro Pro Gly Pro Ser Gly Ala Pro Ile Val Gly Asn Trp Leu Gln
        35                  40                  45

GTC GGC GAC GAC CTC AAC CAC CGC AAC CTG ATG GGC CTG GCC AAG CGG     192
Val Gly Asp Asp Leu Asn His Arg Asn Leu Met Gly Leu Ala Lys Arg
    50                  55                  60

TTC GGC GAG GTG TTC CTC CTC CGC ATG GGC GTC CGC AAC CTG GTG GTC     240
Phe Gly Glu Val Phe Leu Leu Arg Met Gly Val Arg Asn Leu Val Val
65                  70                  75                  80

GTC TCC AGC CCC GAG CTC GCC AAG GAG GTC CTC CAC ACC CAG GGC GTC     288
Val Ser Ser Pro Glu Leu Ala Lys Glu Val Leu His Thr Gln Gly Val
                85                  90                  95

GAG TTC GGC TCC CGC ACC CGC AAC GTC GTC TTC GAC ATC TTC ACC GGC     336
Glu Phe Gly Ser Arg Thr Arg Asn Val Val Phe Asp Ile Phe Thr Gly
            100                 105                 110

AAG GGA CAG GAC ATG GTG TTC ACG GTG TAC GGC GAC CAC TGG CGC AAG     384
Lys Gly Gln Asp Met Val Phe Thr Val Tyr Gly Asp His Trp Arg Lys
        115                 120                 125

ATG CGG CGG ATC ATG ACG GTG CCC TTC TTC ACC AAC AAG GTG GTG GCG     432
Met Arg Arg Ile Met Thr Val Pro Phe Phe Thr Asn Lys Val Val Ala
    130                 135                 140

CAG AAC CGC GTG GGG TGG GAG GAG GAG GCC CGG CTG GTG GTG GAG GAC     480
Gln Asn Arg Val Gly Trp Glu Glu Glu Ala Arg Leu Val Val Glu Asp
145                 150                 155                 160

CTC AAG GCC GAC CCG GCG GCG GCG ACG GCG GGC GTG GTG GTC CGC CGC     528
Leu Lys Ala Asp Pro Ala Ala Ala Thr Ala Gly Val Val Val Arg Arg
                165                 170                 175

AGG CTG CAG CTC ATG ATG TAC AAC GAC ATG TTC CGC ATC ATG TTC GAC     576
Arg Leu Gln Leu Met Met Tyr Asn Asp Met Phe Arg Ile Met Phe Asp
            180                 185                 190

CGC CGG TTC GAG AGC GTG GCC GAC CCG CTC TTC AAC CAG CTC AAG GCG     624
Arg Arg Phe Glu Ser Val Ala Asp Pro Leu Phe Asn Gln Leu Lys Ala
        195                 200                 205

CTC AAC GCC GAG CGC AGC ATC CTC TCC CAG AGC TTC GAC TAC AAC TAC     672
Leu Asn Ala Glu Arg Ser Ile Leu Ser Gln Ser Phe Asp Tyr Asn Tyr
    210                 215                 220

GGC GAC TTC ATC CCC GTC CTC CGC CCC TTC CTC CGC CGC TAC CTC AAC     720
Gly Asp Phe Ile Pro Val Leu Arg Pro Phe Leu Arg Arg Tyr Leu Asn
225                 230                 235                 240
```

```
CGC TGC ACC AAC CTC AAG ACC AAG CGG ATG AAG GTG TTC GAG GAC CAC        768
Arg Cys Thr Asn Leu Lys Thr Lys Arg Met Lys Val Phe Glu Asp His
            245                 250                 255

TTC GTC CAG CAG CGC AAG GAG GCG TTG GAG AAG ACG GGT GAG ATC AGG        816
Phe Val Gln Gln Arg Lys Glu Ala Leu Glu Lys Thr Gly Glu Ile Arg
            260                 265                 270

TGC GCC ATG GAC CAC ATC CTG GAA GCC GAA AGG AAG GGC GAG ATC AAC        864
Cys Ala Met Asp His Ile Leu Glu Ala Glu Arg Lys Gly Glu Ile Asn
        275                 280                 285

CAC GAC AAC GTC CTC TAC ATC GTC GAG AAC ATC AAC GTC GCA GCC ATC        912
His Asp Asn Val Leu Tyr Ile Val Glu Asn Ile Asn Val Ala Ala Ile
    290                 295                 300

GAG ACG ACG CTG TGG TCG ATC GAG TGG GGC CTC GCG GAG CTG GTG AAC        960
Glu Thr Thr Leu Trp Ser Ile Glu Trp Gly Leu Ala Glu Leu Val Asn
305                 310                 315                 320

CAC CCG GAG ATC CAG CAG AAG CTG CGC GAG GAG ATC GTC GCC GTT CTG       1008
His Pro Glu Ile Gln Gln Lys Leu Arg Glu Glu Ile Val Ala Val Leu
            325                 330                 335

GGC GCC GGC GTG GCG GTG ACG GAG CCG GAC CTG GAG CGC CTC CCC TAC       1056
Gly Ala Gly Val Ala Val Thr Glu Pro Asp Leu Glu Arg Leu Pro Tyr
        340                 345                 350

CTG CAG TCC GTG GTG AAG GAG ACG CTC CGC CTC CGC ATG GCA ATC CCG       1104
Leu Gln Ser Val Val Lys Glu Thr Leu Arg Leu Arg Met Ala Ile Pro
    355                 360                 365

CTC CTG GTG CCG CAC ATG AAC CTC AGC GAC GCC AAG CTC GCC GGC TAC       1152
Leu Leu Val Pro His Met Asn Leu Ser Asp Ala Lys Leu Ala Gly Tyr
    370                 375                 380

GAC ATC CCC GCC GAG TCC AAG ATC CTC GTC AAC GCC TGG TTC CTC GCC       1200
Asp Ile Pro Ala Glu Ser Lys Ile Leu Val Asn Ala Trp Phe Leu Ala
385                 390                 395                 400

AAC GAC CCC AAG CGG TGG GTG CGC GCC GAT GAG TTC AGG CCG GAG AGG       1248
Asn Asp Pro Lys Arg Trp Val Arg Ala Asp Glu Phe Arg Pro Glu Arg
            405                 410                 415

TTC CTC GAG GAG GAG AAG GCC GTC GAG GCC CAC GGC AAC GAT TTC CGG       1296
Phe Leu Glu Glu Glu Lys Ala Val Glu Ala His Gly Asn Asp Phe Arg
        420                 425                 430

TTC GTG CCC TTC GGC GTC GGC CGC CGG AGC TGC CCC GGG ATC ATC CTC       1344
Phe Val Pro Phe Gly Val Gly Arg Arg Ser Cys Pro Gly Ile Ile Leu
    435                 440                 445

GCG CTG CCC ATC ATC GGC ATC ACG CTC GGA CGC CTG GTG CAG AAC TTC       1392
Ala Leu Pro Ile Ile Gly Ile Thr Leu Gly Arg Leu Val Gln Asn Phe
    450                 455                 460

CAG CTG CTG CCG CCG CCG GGG CAG GAC AAG ATC GAC ACC ACC GAG AAG       1440
Gln Leu Leu Pro Pro Pro Gly Gln Asp Lys Ile Asp Thr Thr Glu Lys
465                 470                 475                 480

CCC GGG CAG TTT ACC AAC CAG ATC CTC AAG CAC GCC ACC ATT GTC TGC       1488
Pro Gly Gln Phe Thr Asn Gln Ile Leu Lys His Ala Thr Ile Val Cys
            485                 490                 495

AAG CCA CTC GAG GCT TAA                                               1506
Lys Pro Leu Glu Ala
            500
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1506 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1503

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
ATG GAT GTT TTG TTG TTG GAA AAA GCT TTG TTG GGT TTG TTC GCC GCG        48
Met Asp Val Leu Leu Leu Glu Lys Ala Leu Leu Gly Leu Phe Ala Ala
1               5                   10                  15

GCG GTG CTG GCC ATC GCC GTC GCC AAG CTC ACC GGC AAG CGC TTC CGC        96
Ala Val Leu Ala Ile Ala Val Ala Lys Leu Thr Gly Lys Arg Phe Arg
            20                  25                  30

CTC CCC CCT GGC CCC TCC GGC GCC CCC ATC GTC GGC AAC TGG CTG CAG       144
Leu Pro Pro Gly Pro Ser Gly Ala Pro Ile Val Gly Asn Trp Leu Gln
        35                  40                  45

GTC GGC GAC GAC CTC AAC CAC CGC AAC CTG ATG GGC CTG GCC AAG CGG       192
Val Gly Asp Asp Leu Asn His Arg Asn Leu Met Gly Leu Ala Lys Arg
    50                  55                  60

TTC GGC GAG GTG TTC CTC CTC CGC ATG GGC GTC CGC AAC CTG GTG GTC       240
Phe Gly Glu Val Phe Leu Leu Arg Met Gly Val Arg Asn Leu Val Val
65                  70                  75                  80

GTC TCC AGC CCC GAG CTC GCC AAG GAG GTC CTC CAC ACC CAG GGC GTC       288
Val Ser Ser Pro Glu Leu Ala Lys Glu Val Leu His Thr Gln Gly Val
                85                  90                  95

GAG TTC GGC TCC CGC ACC CGC AAC GTC GTC TTC GAC ATC TTC ACC GGC       336
Glu Phe Gly Ser Arg Thr Arg Asn Val Val Phe Asp Ile Phe Thr Gly
            100                 105                 110

AAG GGA CAG GAC ATG GTG TTC ACG GTG TAC GGC GAC CAC TGG CGC AAG       384
Lys Gly Gln Asp Met Val Phe Thr Val Tyr Gly Asp His Trp Arg Lys
        115                 120                 125

ATG CGG CGG ATC ATG ACG GTG CCC TTC TTC ACC AAC AAG GTG GTG GCG       432
Met Arg Arg Ile Met Thr Val Pro Phe Phe Thr Asn Lys Val Val Ala
    130                 135                 140

CAG AAC CGC GTG GGG TGG GAG GAG GAG GCC CGG CTG GTG GTG GAG GAC       480
Gln Asn Arg Val Gly Trp Glu Glu Glu Ala Arg Leu Val Val Glu Asp
145                 150                 155                 160

CTC AAG GCC GAC CCG GCG GCG GCG ACG GCG GGC GTG GTG GTC CGC CGC       528
Leu Lys Ala Asp Pro Ala Ala Ala Thr Ala Gly Val Val Val Arg Arg
                165                 170                 175

AGG CTG CAG CTC ATG ATG TAC AAC GAC ATG TTC CGC ATC ATG TTC GAC       576
Arg Leu Gln Leu Met Met Tyr Asn Asp Met Phe Arg Ile Met Phe Asp
            180                 185                 190

CGC CGG TTC GAG AGC GTG GCC GAC CCG CTC TTC AAC CAG CTC AAG GCG       624
Arg Arg Phe Glu Ser Val Ala Asp Pro Leu Phe Asn Gln Leu Lys Ala
        195                 200                 205
```

```
CTC AAC GCC GAG CGC AGC ATC CTC TCC CAG AGC TTC GAC TAC AAC TAC          672
Leu Asn Ala Glu Arg Ser Ile Leu Ser Gln Ser Phe Asp Tyr Asn Tyr
    210                 215                 220

GGC GAC TTC ATC CCC GTC CTC CGC CCC TTC CTC CGC CGC TAC CTC AAC          720
Gly Asp Phe Ile Pro Val Leu Arg Pro Phe Leu Arg Arg Tyr Leu Asn
225                 230                 235                 240

CGC TGC ACC AAC CTC AAG ACC AAG CGG ATG AAG GTG TTC GAG GAC CAC          768
Arg Cys Thr Asn Leu Lys Thr Lys Arg Met Lys Val Phe Glu Asp His
                245                 250                 255

TTC GTC CAG CAG CGC AAG GAG GCG TTG GAG AAG ACG GGT GAG ATC AGG          816
Phe Val Gln Gln Arg Lys Glu Ala Leu Glu Lys Thr Gly Glu Ile Arg
            260                 265                 270

TGC GCC ATG GAC CAC ATC CTG GAA GCC GAA AGG AAG GGC GAG ATC AAC          864
Cys Ala Met Asp His Ile Leu Glu Ala Glu Arg Lys Gly Glu Ile Asn
        275                 280                 285

CAC GAC AAC GTC CTC TAC ATC GTC GAG AAC ATC AAC GTC GCA GCC ATC          912
His Asp Asn Val Leu Tyr Ile Val Glu Asn Ile Asn Val Ala Ala Ile
    290                 295                 300

GAG ACG ACG CTG TGG TCG ATC GAG TGG GGC CTC GCG GAG CTG GTG AAC          960
Glu Thr Thr Leu Trp Ser Ile Glu Trp Gly Leu Ala Glu Leu Val Asn
305                 310                 315                 320

CAC CCG GAG ATC CAG CAG AAG CTG CGC GAG GAG ATC GTC GCC GTT CTG         1008
His Pro Glu Ile Gln Gln Lys Leu Arg Glu Glu Ile Val Ala Val Leu
                325                 330                 335

GGC GCC GGC GTG GCG GTG ACG GAG CCG GAC CTG GAG CGC CTC CCC TAC         1056
Gly Ala Gly Val Ala Val Thr Glu Pro Asp Leu Glu Arg Leu Pro Tyr
            340                 345                 350

CTG CAG TCC GTG GTG AAG GAG ACG CTC CGC CTC CGC ATG GCA ATC CCG         1104
Leu Gln Ser Val Val Lys Glu Thr Leu Arg Leu Arg Met Ala Ile Pro
        355                 360                 365

CTC CTG GTG CCG CAC ATG AAC CTC AGC GAC GCC AAG CTC GCC GGC TAC         1152
Leu Leu Val Pro His Met Asn Leu Ser Asp Ala Lys Leu Ala Gly Tyr
    370                 375                 380

GAC ATC CCC GCC GAG TCC AAG ATC CTC GTC AAC GCC TGG TTC CTC GCC         1200
Asp Ile Pro Ala Glu Ser Lys Ile Leu Val Asn Ala Trp Phe Leu Ala
385                 390                 395                 400

AAC GAC CCC AAG CGG TGG GTG CGC GCC GAT GAG TTC AGG CCG GAG AGG         1248
Asn Asp Pro Lys Arg Trp Val Arg Ala Asp Glu Phe Arg Pro Glu Arg
                405                 410                 415

TTC CTC GAG GAG GAG AAG GCC GTC GAG GCC CAC GGC AAC GAT TTC CGG         1296
Phe Leu Glu Glu Glu Lys Ala Val Glu Ala His Gly Asn Asp Phe Arg
            420                 425                 430

TTC GTG CCC TTC GGC GTC GGC CGC CGG AGC TGC CCC GGG ATC ATC CTC         1344
Phe Val Pro Phe Gly Val Gly Arg Arg Ser Cys Pro Gly Ile Ile Leu
        435                 440                 445

GCG CTG CCC ATC ATC GGC ATC ACG CTC GGA CGC CTG GTG CAG AAC TTC         1392
Ala Leu Pro Ile Ile Gly Ile Thr Leu Gly Arg Leu Val Gln Asn Phe
    450                 455                 460

CAG CTG CTG CCG CCG CCG GGG CAG GAC AAG ATC GAC ACC ACC GAG AAG         1440
Gln Leu Leu Pro Pro Pro Gly Gln Asp Lys Ile Asp Thr Thr Glu Lys
465                 470                 475                 480
```

```
CCC GGG CAG TTT ACC AAC CAG ATC CTC AAG CAC GCC ACC ATT GTC TGC      1488
Pro Gly Gln Phe Thr Asn Gln Ile Leu Lys His Ala Thr Ile Val Cys
                485                 490                 495

AAG CCA CTC GAG GCT TAA                                              1506
Lys Pro Leu Glu Ala
            500
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1506 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1503

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
ATG GAT GTT TTG TTG TTG GAA AAA GCT TTG TTG GGT TTG TTT GCT GCT        48
Met Asp Val Leu Leu Leu Glu Lys Ala Leu Leu Gly Leu Phe Ala Ala
            505                 510                 515

GCT GTT TTG GCT ATT GCT GTT GCT AAA TTG ACT GGT AAA AGA TTT AGA        96
Ala Val Leu Ala Ile Ala Val Ala Lys Leu Thr Gly Lys Arg Phe Arg
        520                 525                 530

TTG CCA CCA GGT CCA TCC GGC GCC CCC ATC GTC GGC AAC TGG CTG CAG       144
Leu Pro Pro Gly Pro Ser Gly Ala Pro Ile Val Gly Asn Trp Leu Gln
        35                  40                  45

GTC GGC GAC GAC CTC AAC CAC CGC AAC CTG ATG GGC CTG GCC AAG CGG       192
Val Gly Asp Asp Leu Asn His Arg Asn Leu Met Gly Leu Ala Lys Arg
    50                  55                  60

TTC GGC GAG GTG TTC CTC CTC CGC ATG GGC GTC CGC AAC CTG GTG GTC       240
Phe Gly Glu Val Phe Leu Leu Arg Met Gly Val Arg Asn Leu Val Val
65                  70                  75                  80

GTC TCC AGC CCC GAG CTC GCC AAG GAG GTC CTC CAC ACC CAG GGC GTC       288
Val Ser Ser Pro Glu Leu Ala Lys Glu Val Leu His Thr Gln Gly Val
                85                  90                  95

GAG TTC GGC TCC CGC ACC CGC AAC GTC GTC TTC GAC ATC TTC ACC GGC       336
Glu Phe Gly Ser Arg Thr Arg Asn Val Val Phe Asp Ile Phe Thr Gly
            100                 105                 110

AAG GGA CAG GAC ATG GTG TTC ACG GTG TAC GGC GAC CAC TGG CGC AAG       384
Lys Gly Gln Asp Met Val Phe Thr Val Tyr Gly Asp His Trp Arg Lys
        115                 120                 125

ATG CGG CGG ATC ATG ACG GTG CCC TTC TTC ACC AAC AAG GTG GTG GCG       432
Met Arg Arg Ile Met Thr Val Pro Phe Phe Thr Asn Lys Val Val Ala
    130                 135                 140

CAG AAC CGC GTG GGG TGG GAG GAG GAG GCC CGG CTG GTG GTG GAG GAC       480
Gln Asn Arg Val Gly Trp Glu Glu Glu Ala Arg Leu Val Val Glu Asp
145                 150                 155                 160

CTC AAG GCC GAC CCG GCG GCG GCG ACG GCG GGC GTG GTG GTC CGC CGC       528
Leu Lys Ala Asp Pro Ala Ala Ala Thr Ala Gly Val Val Val Arg Arg
                165                 170                 175
```

AGG CTG CAG CTC ATG ATG TAC AAC GAC ATG TTC CGC ATC ATG TTC GAC 576
Arg Leu Gln Leu Met Met Tyr Asn Asp Met Phe Arg Ile Met Phe Asp
180 185 190

CGC CGG TTC GAG AGC GTG GCC GAC CCG CTC TTC AAC CAG CTC AAG GCG 624
Arg Arg Phe Glu Ser Val Ala Asp Pro Leu Phe Asn Gln Leu Lys Ala
195 200 205

CTC AAC GCC GAG CGC AGC ATC CTC TCC CAG AGC TTC GAC TAC AAC TAC 672
Leu Asn Ala Glu Arg Ser Ile Leu Ser Gln Ser Phe Asp Tyr Asn Tyr
210 215 220

GGC GAC TTC ATC CCC GTC CTC CGC CCC TTC CTC CGC CGC TAC CTC AAC 720
Gly Asp Phe Ile Pro Val Leu Arg Pro Phe Leu Arg Arg Tyr Leu Asn
225 230 235 240

CGC TGC ACC AAC CTC AAG ACC AAG CGG ATG AAG GTG TTC GAG GAC CAC 768
Arg Cys Thr Asn Leu Lys Thr Lys Arg Met Lys Val Phe Glu Asp His
245 250 255

TTC GTC CAG CAG CGC AAG GAG GCG TTG GAG AAG ACG GGT GAG ATC AGG 816
Phe Val Gln Gln Arg Lys Glu Ala Leu Glu Lys Thr Gly Glu Ile Arg
260 265 270

TGC GCC ATG GAC CAC ATC CTG GAA GCC GAA AGG AAG GGC GAG ATC AAC 864
Cys Ala Met Asp His Ile Leu Glu Ala Glu Arg Lys Gly Glu Ile Asn
275 280 285

CAC GAC AAC GTC CTC TAC ATC GTC GAG AAC ATC AAC GTC GCA GCC ATC 912
His Asp Asn Val Leu Tyr Ile Val Glu Asn Ile Asn Val Ala Ala Ile
290 295 300

GAG ACG ACG CTG TGG TCG ATC GAG TGG GGC CTC GCG GAG CTG GTG AAC 960
Glu Thr Thr Leu Trp Ser Ile Glu Trp Gly Leu Ala Glu Leu Val Asn
305 310 315 320

CAC CCG GAG ATC CAG CAG AAG CTG CGC GAG GAG ATC GTC GCC GTT CTG 1008
His Pro Glu Ile Gln Gln Lys Leu Arg Glu Glu Ile Val Ala Val Leu
325 330 335

GGC GCC GGC GTG GCG GTG ACG GAG CCG GAC CTG GAG CGC CTC CCC TAC 1056
Gly Ala Gly Val Ala Val Thr Glu Pro Asp Leu Glu Arg Leu Pro Tyr
340 345 350

CTG CAG TCC GTG GTG AAG GAG ACG CTC CGC CTC CGC ATG GCA ATC CCG 1104
Leu Gln Ser Val Val Lys Glu Thr Leu Arg Leu Arg Met Ala Ile Pro
355 360 365

CTC CTG GTG CCG CAC ATG AAC CTC AGC GAC GCC AAG CTC GCC GGC TAC 1152
Leu Leu Val Pro His Met Asn Leu Ser Asp Ala Lys Leu Ala Gly Tyr
370 375 380

GAC ATC CCC GCC GAG TCC AAG ATC CTC GTC AAC GCC TGG TTC CTC GCC 1200
Asp Ile Pro Ala Glu Ser Lys Ile Leu Val Asn Ala Trp Phe Leu Ala
385 390 395 400

AAC GAC CCC AAG CGG TGG GTG CGC GCC GAT GAG TTC AGG CCG GAG AGG 1248
Asn Asp Pro Lys Arg Trp Val Arg Ala Asp Glu Phe Arg Pro Glu Arg
405 410 415

TTC CTC GAG GAG GAG AAG GCC GTC GAG GCC CAC GGC AAC GAT TTC CGG 1296
Phe Leu Glu Glu Glu Lys Ala Val Glu Ala His Gly Asn Asp Phe Arg
420 425 430

TTC GTG CCC TTC GGC GTC GGC CGC CGG AGC TGC CCC GGG ATC ATC CTC 1344
Phe Val Pro Phe Gly Val Gly Arg Arg Ser Cys Pro Gly Ile Ile Leu
435 440 445

```
GCG CTG CCC ATC ATC GGC ATC ACG CTC GGA CGC CTG GTG CAG AAC TTC        1392
Ala Leu Pro Ile Ile Gly Ile Thr Leu Gly Arg Leu Val Gln Asn Phe
    450                 455                 460

CAG CTG CTG CCG CCG CCG GGG CAG GAC AAG ATC GAC ACC ACC GAG AAG        1440
Gln Leu Leu Pro Pro Pro Gly Gln Asp Lys Ile Asp Thr Thr Glu Lys
465                 470                 475                 480

CCC GGG CAG TTT ACC AAC CAG ATC CTC AAG CAC GCC ACC ATT GTC TGC        1488
Pro Gly Gln Phe Thr Asn Gln Ile Leu Lys His Ala Thr Ile Val Cys
            485                 490                 495

AAG CCA CTC GAG GCT TAA                                                1506
Lys Pro Leu Glu Ala
            500
```

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2181 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 112..1734

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
CGATCCACCC CTTGGATCCA CTCTACCCAG CTCGCTAGCC AGCGGGGTAC ATACACGCAC        60

GCACGTACGC GCGTACGTAC ACTCGCAGAG CTTGCTTCAG GGAGGCCGGC A ATG GAG       117
                                                         Met Glu
                                                           1

GTG GGG ACG TGG GCG GTG GTG GTG TCG GCG GTG GCC GCG TAC ATG GCG       165
Val Gly Thr Trp Ala Val Val Val Ser Ala Val Ala Ala Tyr Met Ala
         5                  10                  15

TGG TTC TGG CGG ATG TCC CGC GGG CTG CGC GGG CCG CGG GTT TGG CCC       213
Trp Phe Trp Arg Met Ser Arg Gly Leu Arg Gly Pro Arg Val Trp Pro
     20                  25                  30

GTG CTC GGC AGC CTG CCG GGC CTG GTG CAG CAC GCC GAG GAC ATG CAC       261
Val Leu Gly Ser Leu Pro Gly Leu Val Gln His Ala Glu Asp Met His
 35                  40                  45                  50

GAG TGG ATC GCC GGC AAC CTG CGC CGC GCG GGC GGC ACG TAC CAG ACC       309
Glu Trp Ile Ala Gly Asn Leu Arg Arg Ala Gly Gly Thr Tyr Gln Thr
                 55                  60                  65

TGC ATC TTC GCC GTG CCC GGG GTG GCG CGC CGC GGC GGC CTG GTC ACC       357
Cys Ile Phe Ala Val Pro Gly Val Ala Arg Arg Gly Gly Leu Val Thr
             70                  75                  80

GTC ACC TGC GAC CCG CGC AAC CTG GAG CAC GTC CTG AAG GCG CGC TTC       405
Val Thr Cys Asp Pro Arg Asn Leu Glu His Val Leu Lys Ala Arg Phe
         85                  90                  95

GAC AAC TAC CCC AAG GGC CCC TTC TGG CAC GGC GTC TTC CGG GAC CTG       453
Asp Asn Tyr Pro Lys Gly Pro Phe Trp His Gly Val Phe Arg Asp Leu
    100                 105                 110
```

```
CTC GGC GAC GGC ATC TTC AAT TCC GAC GGC GAC ACC TGG CTC GCG CAG      501
Leu Gly Asp Gly Ile Phe Asn Ser Asp Gly Asp Thr Trp Leu Ala Gln
115                 120                 125                 130

CGC AAG ACG GCC GCG CTC GAG TTC ACC ACC CGC ACG CTC CGG ACG GCC      549
Arg Lys Thr Ala Ala Leu Glu Phe Thr Thr Arg Thr Leu Arg Thr Ala
                135                 140                 145

ATG TCC CGC TGG GTC TCG CGC TCC ATC CAC GGC CGC CTC CTG CCC ATC      597
Met Ser Arg Trp Val Ser Arg Ser Ile His Gly Arg Leu Leu Pro Ile
            150                 155                 160

CTG GCC GAC GCG GCC AAG GGC AAG GCG CAG GTG GAT CTC CAG GAC CTC      645
Leu Ala Asp Ala Ala Lys Gly Lys Ala Gln Val Asp Leu Gln Asp Leu
        165                 170                 175

CTC CTC CGC CTC ACC TTC GAC AAC ATC TGC GGC CTG GCC TTC GGC AAG      693
Leu Leu Arg Leu Thr Phe Asp Asn Ile Cys Gly Leu Ala Phe Gly Lys
    180                 185                 190

GAC CCG GAG ACG CTC GCC CAG GGC CTG CCG GAG AAC GAG TTC GCC TCC      741
Asp Pro Glu Thr Leu Ala Gln Gly Leu Pro Glu Asn Glu Phe Ala Ser
195                 200                 205                 210

GCG TTC GAC CGC GCC ACC GAG GCC ACG CTC AAC CGC TTC ATC TTC CCG      789
Ala Phe Asp Arg Ala Thr Glu Ala Thr Leu Asn Arg Phe Ile Phe Pro
                215                 220                 225

GAG TTC CTG TGG CGC TGC AAA AAG TGG CTG GGC CTC GGC ATG GAG ACC      837
Glu Phe Leu Trp Arg Cys Lys Lys Trp Leu Gly Leu Gly Met Glu Thr
            230                 235                 240

ACG CTG ACC AGC AGC ATG GCC CAC GTC GAC CAG TAC CTC GCC GCC GTC      885
Thr Leu Thr Ser Ser Met Ala His Val Asp Gln Tyr Leu Ala Ala Val
        245                 250                 255

ATC AAG AAG CGC AAG CTC GAG CTC GCC GCC GGC AAC GGC AAA TGC GAC      933
Ile Lys Lys Arg Lys Leu Glu Leu Ala Ala Gly Asn Gly Lys Cys Asp
    260                 265                 270

ACG GCG GCG ACG CAC GAC GAC CTG CTC TCC CGG TTC ATG CGG AAG GGT      981
Thr Ala Ala Thr His Asp Asp Leu Leu Ser Arg Phe Met Arg Lys Gly
275                 280                 285                 290

TCC TAC TCG GAC GAG TCG CTC CAG CAC GTG GCG CTC AAC TTC ATC CTC     1029
Ser Tyr Ser Asp Glu Ser Leu Gln His Val Ala Leu Asn Phe Ile Leu
                295                 300                 305

GCC GGC CGC GAC ACC TCC TCC GTG GCG CTC TCC TGG TTC TTC TGG CTC     1077
Ala Gly Arg Asp Thr Ser Ser Val Ala Leu Ser Trp Phe Phe Trp Leu
            310                 315                 320

GTG TCC ACC CAC CCT GCG GTG GAG CGC AAG ATC GTG CGC GAG CTC TGC     1125
Val Ser Thr His Pro Ala Val Glu Arg Lys Ile Val Arg Glu Leu Cys
        325                 330                 335

TCC GTT CTC GCC GCG TCA CGG GGC GCC CAT GAC CCG GCA TTG TGG CTG     1173
Ser Val Leu Ala Ala Ser Arg Gly Ala His Asp Pro Ala Leu Trp Leu
    340                 345                 350

GCG GAG CCC TTC ACC TTC GAG GAG CTC GAC CGC CTG GTC TAC CTC AAG     1221
Ala Glu Pro Phe Thr Phe Glu Glu Leu Asp Arg Leu Val Tyr Leu Lys
355                 360                 365                 370

GCG GCG CTG TCG GAG ACC CTC CGC CTC TAC CCC TCC GTC CCC GAG GAC     1269
Ala Ala Leu Ser Glu Thr Leu Arg Leu Tyr Pro Ser Val Pro Glu Asp
                375                 380                 385
```

```
TCC AAG CAC GTC GTC GCG GAC GAC TAC CTC CCC GAC GGC ACC TTC GTG    1317
Ser Lys His Val Val Ala Asp Asp Tyr Leu Pro Asp Gly Thr Phe Val
            390                 395                 400

CCG GCC GGG TCG TCG GTC ACC TAC TCC ATA TAC TCG GCG GGG CGC ATG    1365
Pro Ala Gly Ser Ser Val Thr Tyr Ser Ile Tyr Ser Ala Gly Arg Met
        405                 410                 415

AAG GGG GTG TGG GGG GAG GAC TGC CTC GAG TTC CGG CCG GAG CGA TGG    1413
Lys Gly Val Trp Gly Glu Asp Cys Leu Glu Phe Arg Pro Glu Arg Trp
    420                 425                 430

CTG TCG GCC GAC GGC ACC AAG TTC GAG CAG CAC GAC TCG TAC AAG TTC    1461
Leu Ser Ala Asp Gly Thr Lys Phe Glu Gln His Asp Ser Tyr Lys Phe
435                 440                 445                 450

GTG GCG TTC AAC GCC GGG CCG AGG GTG TGC CTG GGC AAG GAC CTA GCC    1509
Val Ala Phe Asn Ala Gly Pro Arg Val Cys Leu Gly Lys Asp Leu Ala
                455                 460                 465

TAC CTG CAG ATG AAG AAC ATC GCC GGG AGC GTG CTG CTC CGG CAC CGC    1557
Tyr Leu Gln Met Lys Asn Ile Ala Gly Ser Val Leu Leu Arg His Arg
            470                 475                 480

CTG ACC GTG GCG CCG GGC CAC CGC GTG GAG CAG AAG ATG TCG CTC ACG    1605
Leu Thr Val Ala Pro Gly His Arg Val Glu Gln Lys Met Ser Leu Thr
        485                 490                 495

CTC TTC ATG AAG GGC GGG CTA CGG ATG GAG GTA CGT CCG CGC GAC CTC    1653
Leu Phe Met Lys Gly Gly Leu Arg Met Glu Val Arg Pro Arg Asp Leu
    500                 505                 510

GCC CCC GTC CTC GAC GAG CCC TGC GGC CTG GAC GCC GGC GCC GCC ACC    1701
Ala Pro Val Leu Asp Glu Pro Cys Gly Leu Asp Ala Gly Ala Ala Thr
515                 520                 525                 530

GCC GCC GCA GCA AGT GCC ACA GCG CCG TGC GCG TAGAAGACCT GGCACCGGCA    1754
Ala Ala Ala Ala Ser Ala Thr Ala Pro Cys Ala
                535                 540

CGCGCCATGC ATGATTCGTG CGTGCTAGCT GTTGAAGGGA CGCCGGACAT TGAATGTGTA    1814

GATAGGGCAG CAGTGCAAGA CCGTAAGTAA AATTGATGAT GGGTTTGGTG ACAACATTGA    1874

AGCCACTCCT TTCCAGAATT TACGACCCGG ATAGGAGAAA CAGGGAAACT TTGCAGATCA    1934

CAACACAAGA TCTAGCCAGC CGGGGATCTG ATCTGATTTG CGTCTGCTCG GAGCACGGGT    1994

GCATGGGAGA CCAAGGAGGA AAACAAAAAA TAACAGAAAC AGAGTGAGCA ATATTTGTGA    2054

TTGTAGCCAC GGGAAAGAGA GAGGAGTAAT TAGTAATTCA GATTTGTTTG CAGTAGCTCG    2114

GTGTTGGTGA CCAGATCATA GCCAACTAGG CTATTCTATT CTATTCTATT TTTGAAGATG    2174

ATTTTTC                                                            2181
```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 150 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

**ATATATGGAT CCATGGAGGT GGGGACGTGG GCGGTGGTG** 39

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 150 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

**ATATATGGAT CCATGGAAGT TGGTACTTGG GCTGTTGTTG TTTCTGCTGT TGCTGCTTAT** 60
**ATGGCTTGGT TTTGGAGAAT GTCTAGAGGT TTGAGAGGTC CAAGAGTTTG GCCAGTTTTG** 120
**GGTTCTTTGC CAGGCCTGGT GCAGCACGCC** 150

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "reverse"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

```
TATATAGAAT TCCTTCTACG CGCACGGCGC TGTGGCACTT GC                          42
```

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1626 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1623

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
ATG GAA GTT GGT ACT TGG GCT GTT GTT GTT TCT GCT GTT GCT GCT TAT       48
Met Glu Val Gly Thr Trp Ala Val Val Val Ser Ala Val Ala Ala Tyr
  1               5                  10                  15

ATG GCT TGG TTT TGG AGA ATG TCT AGA GGT TTG AGA GGT CCA AGA GTT       96
Met Ala Trp Phe Trp Arg Met Ser Arg Gly Leu Arg Gly Pro Arg Val
             20                  25                  30

TGG CCA GTT TTG GGT TCT TTG CCA GGC CTG GTG CAG CAC GCC GAG GAC      144
Trp Pro Val Leu Gly Ser Leu Pro Gly Leu Val Gln His Ala Glu Asp
         35                  40                  45
```

```
ATG CAC GAG TGG ATC GCC GGC AAC CTG CGC CGC GCG GGC GGC ACG TAC      192
Met His Glu Trp Ile Ala Gly Asn Leu Arg Arg Ala Gly Gly Thr Tyr
     50                  55                  60

CAG ACC TGC ATC TTC GCC GTG CCC GGG GTG GCG CGC CGC GGC GGC CTG      240
Gln Thr Cys Ile Phe Ala Val Pro Gly Val Ala Arg Arg Gly Gly Leu
 65                  70                  75                  80

GTC ACC GTC ACC TGC GAC CCG CGC AAC CTG GAG CAC GTC CTG AAG GCG      288
Val Thr Val Thr Cys Asp Pro Arg Asn Leu Glu His Val Leu Lys Ala
                 85                  90                  95

CGC TTC GAC AAC TAC CCC AAG GGC CCC TTC TGG CAC GGC GTC TTC CGG      336
Arg Phe Asp Asn Tyr Pro Lys Gly Pro Phe Trp His Gly Val Phe Arg
            100                 105                 110

GAC CTG CTC GGC GAC GGC ATC TTC AAT TCC GAC GGC GAC ACC TGG CTC      384
Asp Leu Leu Gly Asp Gly Ile Phe Asn Ser Asp Gly Asp Thr Trp Leu
        115                 120                 125

GCG CAG CGC AAG ACG GCC GCG CTC GAG TTC ACC ACC CGC ACG CTC CGG      432
Ala Gln Arg Lys Thr Ala Ala Leu Glu Phe Thr Thr Arg Thr Leu Arg
    130                 135                 140

ACG GCC ATG TCC CGC TGG GTC TCG CGC TCC ATC CAC GGC CGC CTC CTG      480
Thr Ala Met Ser Arg Trp Val Ser Arg Ser Ile His Gly Arg Leu Leu
145                 150                 155                 160

CCC ATC CTG GCC GAC GCG GCC AAG GGC AAG GCG CAG GTG GAT CTC CAG      528
Pro Ile Leu Ala Asp Ala Ala Lys Gly Lys Ala Gln Val Asp Leu Gln
                165                 170                 175

GAC CTC CTC CTC CGC CTC ACC TTC GAC AAC ATC TGC GGC CTG GCC TTC      576
Asp Leu Leu Leu Arg Leu Thr Phe Asp Asn Ile Cys Gly Leu Ala Phe
            180                 185                 190

GGC AAG GAC CCG GAG ACG CTC GCC CAG GGC CTG CCG GAG AAC GAG TTC      624
Gly Lys Asp Pro Glu Thr Leu Ala Gln Gly Leu Pro Glu Asn Glu Phe
        195                 200                 205

GCC TCC GCG TTC GAC CGC GCC ACC GAG GCC ACG CTC AAC CGC TTC ATC      672
Ala Ser Ala Phe Asp Arg Ala Thr Glu Ala Thr Leu Asn Arg Phe Ile
    210                 215                 220

TTC CCG GAG TTC CTG TGG CGC TGC AAA AAG TGG CTG GGC CTC GGC ATG      720
Phe Pro Glu Phe Leu Trp Arg Cys Lys Lys Trp Leu Gly Leu Gly Met
225                 230                 235                 240

GAG ACC ACG CTG ACC AGC AGC ATG GCC CAC GTC GAC CAG TAC CTC GCC      768
Glu Thr Thr Leu Thr Ser Ser Met Ala His Val Asp Gln Tyr Leu Ala
                245                 250                 255

GCC GTC ATC AAG AAG CGC AAG CTC GAG CTC GCC GCC GGC AAC GGC AAA      816
Ala Val Ile Lys Lys Arg Lys Leu Glu Leu Ala Ala Gly Asn Gly Lys
            260                 265                 270

TGC GAC ACG GCG GCG ACG CAC GAC GAC CTG CTC TCC CGG TTC ATG CGG      864
Cys Asp Thr Ala Ala Thr His Asp Asp Leu Leu Ser Arg Phe Met Arg
        275                 280                 285

AAG GGT TCC TAC TCG GAC GAG TCG CTC CAG CAC GTG GCG CTC AAC TTC      912
Lys Gly Ser Tyr Ser Asp Glu Ser Leu Gln His Val Ala Leu Asn Phe
    290                 295                 300

ATC CTC GCC GGC CGC GAC ACC TCC TCC GTG GCG CTC TCC TGG TTC TTC      960
Ile Leu Ala Gly Arg Asp Thr Ser Ser Val Ala Leu Ser Trp Phe Phe
305                 310                 315                 320
```

```
TGG CTC GTG TCC ACC CAC CCT GCG GTG GAG CGC AAG ATC GTG CGC GAG      1008
Trp Leu Val Ser Thr His Pro Ala Val Glu Arg Lys Ile Val Arg Glu
                325                 330                 335

CTC TGC TCC GTT CTC GCC GCG TCA CGG GGC GCC CAT GAC CCG GCA TTG      1056
Leu Cys Ser Val Leu Ala Ala Ser Arg Gly Ala His Asp Pro Ala Leu
            340                 345                 350

TGG CTG GCG GAG CCC TTC ACC TTC GAG GAG CTC GAC CGC CTG GTC TAC      1104
Trp Leu Ala Glu Pro Phe Thr Phe Glu Glu Leu Asp Arg Leu Val Tyr
        355                 360                 365

CTC AAG GCG GCG CTG TCG GAG ACC CTC CGC CTC TAC CCC TCC GTC CCC      1152
Leu Lys Ala Ala Leu Ser Glu Thr Leu Arg Leu Tyr Pro Ser Val Pro
    370                 375                 380

GAG GAC TCC AAG CAC GTC GTC GCG GAC GAC TAC CTC CCC GAC GGC ACC      1200
Glu Asp Ser Lys His Val Val Ala Asp Asp Tyr Leu Pro Asp Gly Thr
385                 390                 395                 400

TTC GTG CCG GCC GGG TCG TCG GTC ACC TAC TCC ATA TAC TCG GCG GGG      1248
Phe Val Pro Ala Gly Ser Ser Val Thr Tyr Ser Ile Tyr Ser Ala Gly
                405                 410                 415

CGC ATG AAG GGG GTG TGG GGG GAG GAC TGC CTC GAG TTC CGG CCG GAG      1296
Arg Met Lys Gly Val Trp Gly Glu Asp Cys Leu Glu Phe Arg Pro Glu
            420                 425                 430

CGA TGG CTG TCG GCC GAC GGC ACC AAG TTC GAG CAG CAC GAC TCG TAC      1344
Arg Trp Leu Ser Ala Asp Gly Thr Lys Phe Glu Gln His Asp Ser Tyr
        435                 440                 445

AAG TTC GTG GCG TTC AAC GCC GGG CCG AGG GTG TGC CTG GGC AAG GAC      1392
Lys Phe Val Ala Phe Asn Ala Gly Pro Arg Val Cys Leu Gly Lys Asp
    450                 455                 460

CTA GCC TAC CTG CAG ATG AAG AAC ATC GCC GGG AGC GTG CTG CTC CGG      1440
Leu Ala Tyr Leu Gln Met Lys Asn Ile Ala Gly Ser Val Leu Leu Arg
465                 470                 475                 480

CAC CGC CTG ACC GTG GCG CCG GGC CAC CGC GTG GAG CAG AAG ATG TCG      1488
His Arg Leu Thr Val Ala Pro Gly His Arg Val Glu Gln Lys Met Ser
                485                 490                 495

CTC ACG CTC TTC ATG AAG GGC GGG CTA CGG ATG GAG GTA CGT CCG CGC      1536
Leu Thr Leu Phe Met Lys Gly Gly Leu Arg Met Glu Val Arg Pro Arg
            500                 505                 510

GAC CTC GCC CCC GTC CTC GAC GAG CCC TGC GGC CTG GAC GCC GGC GCC      1584
Asp Leu Ala Pro Val Leu Asp Glu Pro Cys Gly Leu Asp Ala Gly Ala
        515                 520                 525

GCC ACC GCC GCC GCA GCA AGT GCC ACA GCG CCG TGC GCG TAG              1626
Ala Thr Ala Ala Ala Ala Ser Ala Thr Ala Pro Cys Ala
    530                 535                 540
```

## CLAIMS

1. DNA sequence which encodes a protein of interest which contains regions having a high content of codons which are poorly suited to yeasts, characterized in that a sufficient number of codons which are poorly suited to yeasts is replaced with corresponding codons which are well-suited to yeasts in the said regions having a high content of codons which are poorly suited to yeasts.

2. Sequence according to claim 1, characterized in that the codons which are poorly suited to yeasts are selected from among codons whose frequency of use by yeasts is less than or equal to approximately 13 per 1000, preferably less than or equal to approximately 12 per 1000, more preferably less than or equal to approximately 10 per 1000.

3. Sequence according to claim 2, characterized in that the codons which are poorly suited to yeasts are selected from among codons CTC, CTG and CTT, which encode leucine, codons CGG, CGC, CGA, CGT and AGG, which encode arginine, codons GCG and GCC, which encode alanine, codons GGG, GGC and GGA, which encode glycine, and codons CCG and CCC, which encode proline.

4. Sequence according to claim 3, characterized in that the codons which are poorly suited to yeasts are selected from among codons CTC and CTG, which encode leucine, codons CGG, CGC, CGA, CGT

and AGG, which encode arginine, codons GCG and GCC, which encode alanine, codons GGG and GGC, which encode glycine, and codons CCG and CCC, which encode proline.

5. Sequence according to one of claims 1 to 4, characterized in that the corresponding codons which are well-suited to yeasts are selected from among codons which correspond to the codons which are poorly suited to yeasts and which encode the same amino acids, and whose frequency of use by yeasts is greater than 15 per 1000, preferably greater than or equal to 18 per 1000, more preferably greater than or equal to 20 per 1000.

6. Sequence according to claim 5, characterized in that the corresponding codons which are well-suited to yeasts are selected from among codons TTG and TTA, preferably TTG, which encode leucine, codon AGA, which encodes arginine, codons GCT and GCA, preferably GCT, which encode alanine, codon GGT, which encodes glycine, and codon CCA, which encodes proline.

7. Sequence according to one of claims 1 to 7, characterized in that the regions having a high content of codons which are poorly suited to yeasts contain at least 2 poorly suited codons among 10 consecutive codons, with it being possible for the two codons to be adjacent or separated by up to 8 other codons.

8. Sequence according to claim 7,

characterized in that the regions having a high content of poorly suited codons contain 2, 3, 4, 5 or 6 poorly suited codons per 10 consecutive codons, or contain at least 2 or 3 adjacent poorly suited codons.

9. DNA, in particular cDNA, sequence which encodes a protein of interest which contains regions having a high content of leucine, characterized in that a sufficient number of CTC codons encoding leucine in the said region having a high content of leucine is replaced with TTG and/or TTA codons, or in that a sufficient number of CTC and CTG codons encoding leucine in the said region having a high content of leucine is replaced with TTG and/or TTA codons.

10. Sequence according to claim 9, characterized in that the CTC or CTC and CTG codons are replaced with a TTG codon.

11. Sequence according to one of claims 9 or 10, characterized in that the regions having a high content of leucine contain 2, 3, 4, 5 or 6 leucines per 10 consecutive amino acids, or contain at least 2 or 3 adjacent leucines.

12. Sequence according to one of claims 1 to 11, characterized in that the general content of poorly suited codons is at least 20%, more preferably at least 30%, as compared with the total number of codons.

13. Sequence according to one of claims 1 to 12, characterized in that it contains at least one 5' region having a high content of codons which are poorly

suited to yeasts.

14. Sequence according to claim 13, characterized in that the codons which are poorly suited to yeasts are replaced only in this 5' region.

15. Sequence according to one of claims 1 to 14, characterized in that it is an isolated DNA sequence of natural origin, in particular of plant origin.

16. Sequence according to claim 15, characterized in that it originates from dicotyledonous or monocotyledonous plants, in particular from monocotyledonous plants.

17. Sequence according to claim 16, characterized in that it originates from plants of the graminae family, which are selected, in particular, from among wheat, barley, oats, rice, maize, sorghum and cane sugar.

18. Sequence according to one of claims 1 to 17, characterized in that it encodes an enzyme.

19. Sequence according to claim 18, characterized in that it encodes a cytochrome P450.

20. Sequence according to claim 19, characterized in that the sequence which contains regions having a high content of codons which are poorly suited to yeasts includes the coding region of the sequences ID No. 1 or ID No. 10.

21. Sequence according to claim 19, characterized in that it is one of the sequences ID

No. 7, ID No. 8, ID No. 9 and ID No. 13.

22. Chimeric gene which contains a modified DNA sequence according to one of claims 1 to 21 and heterologous 5′ and 3′ regulatory elements which are able to function in a yeast.

23. Vector for transforming yeasts which contains at least one chimeric gene according to claim 22.

24. Process for transforming yeasts using a vector according to claim 23.

25. Transformed yeast for expressing a protein of interest, characterized in that it contains a chimeric gene according to claim 22.

26. Yeast according to claim 25, characterized in that it is selected from among the genera *Saccharomyces, Kluyveromyces, Hansenula, Pichia* and *Yarrowia*, advantageously from the genus *Saccharomyces*, in particular *S. cerevisiae*.

27. Process for producing a heterologous protein of interest in a transformed yeast, characterized in that it comprises the steps of:

a) transforming a yeast with a vector according to claim 23 which contains a modified DNA sequence according to one of claims 1 to 21 and heterologous 5′ and 3′ regulatory elements which are able to function in a yeast,

b) culturing the transformed yeast, and

c) extracting the protein of interest from

the yeast culture.

28. Process for transforming a substrate by enzymic catalysis using an enzyme which is expressed in a yeast, which process comprises the steps of

a) culturing, in the presence of the substrate to be transformed, the yeast which has been transformed with a vector according to claim 23 which contains a modified DNA sequence according to one of claims 1 to 21 and heterologous 5′ and 3′ regulatory elements which are able to function in a yeast, and then

b) recovering the transformed substrate from the yeast culture.

RHONE-POULENC AGROCHIMIE

# THE RECODING OF DNA SEQUENCES TO ENABLE THEM TO BE EXPRESSED IN YEASTS, AND THE TRANSFORMED YEASTS OBTAINED

**Abstract**

The present invention relates to a DNA sequence which encodes a protein of interest which contains regions having a high content of codons which are poorly suited to yeasts, characterized in that a sufficient number of codons which are poorly suited to yeasts is replaced with corresponding codons which are well-suited to yeasts in the said regions having a high content of codons which are poorly suited to yeasts.

The present invention relates, more specifically, to DNA sequences which originate from dicotyledonous or monocotyledonous plants, in particular plants of the graminae family which are selected, in particular, from among wheat, barley, oats, rice, maize, sorghum and cane sugar.

The present invention also relates to transformed yeasts which contain a DNA sequence according to the invention.

**COMBINED DECLARATION**
**AND POWER OF ATTORNEY**

**(Original, Design, National Stage of PCT, Divisional, Continuation or C-I-P Application)**

As a below named inventor, I hereby declare that:WE, YANNICK BATARD, ET AL.
My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

RECODING OF DNA SEQUENCES PERMITTING EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST this declaration is of the following type:

[X] original
[] design
[] national stage of PCT.
[] divisional
[] continuation
[] continuation-in-part (C-I-P)

the specification of which: *(complete (a), (b), or (c))*

(a) [ ] is attached hereto.
(b) [X ] was filed on September 23, 1998 as Application Serial No. 09/158,767 and was amended on *(if applicable).*
(c) [ ] was described and claimed in PCT International Application No. filed on and was amended on *(if applicable).*

**Acknowledgement of Review of Papers and Duty of Candor**

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the patentability of the subject matter claimed in this application in accordance with Title 37, Code of Federal Regulations § 1.56.

[ ] In compliance with this duty there is attached an information disclosure statement. 37 CFR 1.98.

**Priority Claim**

I hereby claim foreign priority benefits under Title 35, United States Code, § 119(a)-(d) of any foreign application(s) for patent or inventor's certificate or of any PCT International Application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT International Application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application on which priority is claimed

*(complete (d) or (e))*

(d) [ ] no such applications have been filed.
(e) [X ] such applications have been filed as follows:

| PRIOR FOREIGN/PCT APPLICATION(S) FILED WITHIN 12 MONTHS (6 MONTHS FOR DESIGN) PRIOR TO SAID APPLICATION | | | | |
|---|---|---|---|---|
| COUNTRY | APPLICATION NO. | DATE OF FILING (day, month, year) | DATE OF ISSUE (day, month, year) | PRIORITY CLAIMED UNDER 35 USC 119 |
| FRANCE | 97 12094 | 24-9-97 | | [x ] YES NO [ ] |
| | | | | [ ] YES NO [ ] |
| | | | | [ ] YES NO [ ] |
| ALL FOREIGN APPLICATION[S], IF ANY, FILED MORE THAN 12 MONTHS (6 MONTHS FOR DESIGN) PRIOR TO SAID APPLICATION | | | | |
| | | | | [ ] YES NO [ ] |
| | | | | [ ] YES NO [ ] |
| | | | | [ ] YES NO [ ] |

## Claim for Benefit of Prior U.S. Provisional Application(s)

I hereby claim the benefit under Title 35, United States Code, § 119(e) of any United States provisional application(s) listed below:

| Provisional Application Number | Filing Date |
|---|---|
| | |
| | |
| | |

## Claim for Benefit of Earlier U.S./PCT Application(s) under 35 U.S.C. 120

*(complete this part only if this is a divisional, continuation or C-I-P application)*

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior application(s) in the manner provided by the first paragraph of Title 35, United States Code § 112, I acknowledge the duty to disclose information as defined in Title 37, Code of Federal Regulations, § 1.56 which occurred between the filing date of the prior application(s) and the national or PCT international filing date of this application:

(Application Serial No.) (Filing Date) (Status) (patented, pending, abandoned)

(Application Serial No.) (Filing Date) (Status) (patented, pending, abandoned)

## Power of Attorney

As a named inventor, I hereby appoint Dana M. Raymond, Reg. No. 18,540; Frederick C. Carver, Reg. No. 17,021; Francis J. Hone, Reg. No. 18,662; Joseph D. Garon, Reg. No. 20,420; Arthur S. Tenser, Reg. No. 18,839; Ronald B. Hildreth, Reg. No. 19,498; Thomas R. Nesbitt, Jr., Reg. No. 22,075; Robert Neuner, Reg. No. 24,316; Richard G. Berkley, Reg. No. 25,465; Richard S. Clark, Reg. No. 26,154; Bradley B. Geist, Reg. No. 27,551; James J. Maune, Reg. No. 26,946; John D. Murnane, Reg. No. 29,836, Henry Tang, Reg. No. 29,705, Robert C. Scheinfeld, Reg. No. 31,300, John A. Fogarty, Jr., Reg. No. 22,348, Louis S. Sorell, Reg. No. 32,439 and Rochelle K. Seide Reg. No. 32,300 of the firm of BAKER & BOTTS, L.L.P., with offices at 30 Rockefeller Plaza, New York, New York 10112, as attorneys to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith

| SEND CORRESPONDENCE TO: | DIRECT TELEPHONE CALLS TO: |
|---|---|
| BAKER & BOTTS, L.L.P.<br>30 ROCKEFELLER PLAZA, NEW YORK, N.Y. 10112<br>CUSTOMER NUMBER: 21003 | BAKER & BOTTS, L.L.P.<br>(212) 705-5000 |

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section

001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

| | | | | |
|---|---|---|---|---|
| FULL NAME OF SOLE OR FIRST INVENTOR | LAST NAME<br>BATARD | FIRST NAME<br>YANNICK | MIDDLE NAME | |
| RESIDENCE & CITIZENSHIP | CITY<br>STRASBOURG | STATE or FOREIGN COUNTRY<br>FRANCE | COUNTRY OF CITIZENSHIP<br>FRANCE | |
| POST OFFICE ADDRESS | POST OFFICE ADDRESS<br>5, Rue de l'Aimant | CITY<br>STRASBOURG | STATE or COUNTRY<br>FRANCE | ZIP CODE<br>67000 |
| DATE<br>19/01/99 | SIGNATURE OF INVENTOR<br>Yannick BATARD | | | |
| FULL NAME OF SECOND JOINT INVENTOR, IF ANY | LAST NAME<br>DURST | FIRST NAME<br>FRANCIS | MIDDLE NAME | |
| RESIDENCE & CITIZENSHIP | CITY<br>BERNOLSHEIM | STATE or FOREIGN COUNTRY<br>FRANCE | COUNTRY OF CITIZENSHIP<br>FRANCE | |
| POST OFFICE ADDRESS | POST OFFICE ADDRESS<br>7, Rue de l'Ancienne Ecole | CITY<br>BERNOLSHEIM | STATE or COUNTRY<br>FRANCE | ZIP CODE<br>67170 |
| DATE<br>01/20/99 | SIGNATURE OF INVENTOR<br>Francis DURST | | | |
| FULL NAME OF THIRD JOINT INVENTOR, IF ANY | LAST NAME<br>SCHALK | FIRST NAME<br>MICHEL | MIDDLE NAME | |
| RESIDENCE & CITIZENSHIP | CITY<br>HUTTEHEIM | STATE or FOREIGN COUNTRY<br>FRANCE | COUNTRY OF CITIZENSHIP<br>FRANCE | |
| POST OFFICE ADDRESS | POST OFFICE ADDRESS<br>2, Rue de l'Ungersberg | CITY<br>HUTTEHEIM | STATE or COUNTRY<br>FRANCE | ZIP CODE<br>67230 |
| DATE<br>12.02.99 | SIGNATURE OF INVENTOR | | | |
| FULL NAME OF FOURTH JOINT INVENTOR, IF ANY | LAST NAME<br>WERCK-REICHHART | FIRST NAME<br>DANIELE | MIDDLE NAME | |
| RESIDENCE & CITIZENSHIP | CITY<br>DINGSHEIM | STATE or FOREIGN COUNTRY<br>FRANCE | COUNTRY OF CITIZENSHIP<br>FRANCE | |
| POST OFFICE ADDRESS | POST OFFICE ADDRESS<br>3, Rue de Bagdad | CITY<br>DUNGSHEIM | STATE or COUNTRY<br>FRANCE | ZIP CODE<br>67370 |
| DATE<br>01/22/99 | SIGNATURE OF INVENTOR<br>Daniele WERCK-REICHHART | | | |
| FULL NAME OF FIFTH JOINT INVENTOR, IF ANY | LAST NAME | FIRST NAME | MIDDLE NAME | |
| RESIDENCE & CITIZENSHIP | CITY | STATE or FOREIGN COUNTRY | COUNTRY OF CITIZENSHIP | |
| POST OFFICE ADDRESS | POST OFFICE ADDRESS | CITY | STATE or COUNTRY | ZIP CODE |
| DATE | SIGNATURE OF INVENTOR | | | |
| FULL NAME OF SIXTH JOINT INVENTOR, IF ANY | LAST NAME | FIRST NAME | MIDDLE NAME | |
| RESIDENCE & CITIZENSHIP | CITY | STATE or FOREIGN COUNTRY | COUNTRY OF CITIZENSHIP | |
| POST OFFICE ADDRESS | POST OFFICE ADDRESS | CITY | STATE or COUNTRY | ZIP CODE |
| DATE | SIGNATURE OF INVENTOR | | | |

SEQUENCE LISTING

<110> Batard, Yannick
Durst, Francis
Schalk, Michel
Werck-Reichhart, Daniele

<120> RECODING OF DNA SEQUENCES PERMITTING
EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST

<130> A32000

<140> 09/158,767
<141> 1998-09-23

<150> FR 97-12094
<151> 1997-09-24

<160> 20

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 2261
<212> DNA
<213> Triticum aestivum

<400> 1
```
cgcagcacgg caacacatac acaggagcca cacaccgcac ctaccccgat ggacgtcctc      60
ctcctggaga aggccctcct gggcctcttc gccgcggcgg tgctggccat cgccgtcgcc     120
aagctcaccg gcaagcgctt ccgcctcccc cctggcccct ccggcgcccc catcgtcggc     180
aactggctgc aggtcggcga cgacctcaac caccgcaacc tgatgggcct ggccaagcgg     240
ttcggcgagg tgttcctcct ccgcatgggc gtccgcaacc tggtggtcgt ctccagcccc     300
gagctcgcca aggaggtcct ccacacccag ggcgtcgagt tcggctcccg cacccgcaac     360
gtcgtcttcg acatcttcac cggcaaggga caggacatgg tgttcacggt gtacggcgac     420
cactggcgca agatgcggcg gatcatgacg gtgcccttct tcaccaacaa ggtggtggcg     480
cagaaccgcg tggggtggga ggaggaggcc cggctggtgg tggaggacct caaggccgac     540
ccggcggcgg cgacggcggg cgtggtggtc cgccgcaggc tgcagctcat gatgtacaac     600
gacatgttcc gcatcatgtt cgaccgccgg ttcgagagcg tggccgaccc gctcttcaac     660
cagctcaagg cgctcaacgc cgagcgcagc atcctctccc agagcttcga ctacaactac     720
ggcgacttca tccccgtcct ccgcccccttc ctccgccgct acctcaaccg ctgcaccaac     780
ctcaagacca agcggatgaa ggtgttcgag gaccacttcg tccagcagcg caaggaggcg     840
ttggagaaga cgggtgagat caggtgcgcc atggaccaca tcctggaagc cgaaaggaag     900
```

ggcgagatca accacgacaa cgtcctctac atcgtcgaga acatcaacgt cgcagccatc 960
gagacgacgc tgtggtcgat cgagtggggc ctcgcggagc tggtgaacca cccggagatc 1020
cagcagaagc tgcgcgagga gatcgtcgcc gttctgggcg ccggcgtggc ggtgacggag 1080
ccggacctgg agcgcctccc ctacctgcag tccgtggtga aggagacgct ccgcctccgc 1140
atggcaatcc cgctcctggt gccgcacatg aacctcagcg acgccaagct cgccggctac 1200
gacatccccg ccgagtccaa gatcctcgtc aacgcctggt tcctcgccaa cgaccccaag 1260
cggtgggtgc gcgccgatga gttcaggccg gagaggttcc tcgaggagga gaaggccgtc 1320
gaggcccacg gcaacgattt ccggttcgtg cccttcggcg tcggccgccg gagctgcccc 1380
gggatcatcc tcgcgctgcc catcatcggc atcacgctcg gacgcctggt gcagaacttc 1440
cagctgctgc cgccgccggg gcaggacaag atcgacacca ccgagaagcc cgggcagttt 1500
accaaccaga tcctcaagca cgccaccatt gtctgcaagc cactcgaggc ttaactgaat 1560
tgaggtttcg gtcatgggcg cccgctgacg cggggagatg gatctatgca tgtgactgtg 1620
tattttgcct tctttctttt tggtgttgtt ttttgcagta gtaagtttaa tttttctttg 1680
gtgttggcct atttgtcttc atgtgaggcg tcgtgttgta aatttccata tagttggcaa 1740
tgtgatgtaa aacttggctc caaaaaaaaa aaaaaaaaac tcgagactct tctctctctc 1800
tctctctctc cagcctcggg tctctgctgg caagggaact tgcattaccc tgtgtacgac 1860
ggcgccatgt tcgtccctga agcaccctcc ctgcagagct cccaggacaa cttcgctgca 1920
tctgctggtt tcaagcgtcg aaggagagag ttttgaatac ccgaaagaat atagcgttgg 1980
acatatctgt caaacagggg atcttgctgt gggtctcttg gtgggccaaa tcgcatagac 2040
aatcattcaa atggatgggt tcttcgctgg tcggtcaaaa agtatatgtt gtaattgtac 2100
ggcttttttg ggtcttgttg ccaaagatca tggttattga gttgtgagct ctgagataac 2160
aggtttgtgt atagtgaaat aaagaggagc gtcgtcaaca ccatgtacta tataggcttt 2220
gaaattccat taagatgcat cagaaatcaa tgttggattt g 2261

<210> 2
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic primer

<400> 2
atatatggat ccatggacgt cctcctcctg gagaaggc 38

<210> 3
<211> 56
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic primer

<400> 3
atatatggat ccatggatgt tttgttgttg gagaaggccc tcctgggcct cttcgc 56

<210> 4
<211> 71
<212> DNA
<213> Triticum aestivum

<220>
<223> Synthetic primer

<400> 4
atatatggat ccatggatgt tttgttgttg gaaaaagctt tgttgggttt gttcgccgcg 60
gcggtgctgg c 71

<210> 5
<211> 143
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic primer

<400> 5
atatatggat ccatggatgt tttgttgttg gaaaaagctt tgttgggttt gtttgctgct 60
gctgttttgg ctattgctgt tgctaaattg actggtaaaa gatttagatt gccaccaggt 120
ccatccggcg cccccatcgt cgg 143

<210> 6
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic primer

<400> 6
tatatagaat tccagttaag cctcgagtgg cttgcagac 39

<210> 7
<211> 1506
<212> DNA
<213> Artificial Sequence

<220>
<223> Altered sequences

<400> 7

atggatgttt tgttgttgga gaaggccctc ctgggcctct tcgccgcggc ggtgctggcc 60
atcgccgtcg ccaagctcac cggcaagcgc ttccgcctcc cccctggccc ctccggcgcc 120
cccatcgtcg gcaactggct gcaggtcggc gacgacctca accaccgcaa cctgatgggc 180
ctggccaagc ggttcggcga ggtgttcctc ctccgcatgg gcgtccgcaa cctggtggtc 240
gtctccagcc ccgagctcgc caaggaggtc ctccacaccc agggcgtcga gttcggctcc 300
cgcacccgca acgtcgtctt cgacatcttc accggcaagg gacaggacat ggtgttcacg 360
gtgtacggcg accactggcg caagatgcgg cggatcatga cggtgccctt cttcaccaac 420
aaggtggtgg cgcagaaccg cgtggggtgg gaggaggagg cccggctggt ggtggaggac 480
ctcaaggccg acccggcggc ggcgacggcg ggcgtggtgg tccgccgcag gctgcagctc 540
atgatgtaca acgacatgtt ccgcatcatg ttcgaccgcc ggttcgagag cgtggccgac 600
ccgctcttca accagctcaa ggcgctcaac gccgagcgca gcatcctctc ccagagcttc 660
gactacaact acggcgactt catccccgtc ctccgcccct tcctccgccg ctacctcaac 720
cgctgcacca acctcaagac caagcggatg aaggtgttcg aggaccactt cgtccagcag 780
cgcaaggagg cgttggagaa gacgggtgag atcaggtgcg ccatggacca catcctggaa 840
gccgaaagga agggcgagat caaccacgac aacgtcctct acatcgtcga gaacatcaac 900
gtcgcagcca tcgagacgac gctgtggtcg atcgagtggg gcctcgcgga gctggtgaac 960
cacccggaga tccagcagaa gctgcgcgag gagatcgtcg ccgttctggg cgccggcgtg 1020
gcggtgacgg agccggacct ggagcgcctc ccctacctgc agtccgtggt gaaggagacg 1080
ctccgcctcc gcatggcaat cccgctcctg gtgccgcaca tgaacctcag cgacgccaag 1140
ctcgccggct acgacatccc cgccgagtcc aagatcctcg tcaacgcctg gttcctcgcc 1200
aacgacccca agcggtgggt gcgcgccgat gagttcaggc cggagaggtt cctcgaggag 1260
gagaaggccg tcgaggccca cggcaacgat ttccggttcg tgcccttcgg cgtcggccgc 1320
cggagctgcc ccgggatcat cctcgcgctg cccatcatcg gcatcacgct cggacgcctg 1380
gtgcagaact tccagctgct gccgccgccg ggcaggaca agatcgacac caccgagaag 1440
cccgggcagt ttaccaacca gatcctcaag cacgccacca ttgtctgcaa gccactcgag 1500
gcttaa 1506

<210> 8
<211> 1506
<212> DNA
<213> Artificial Sequence

<220>
<223> Altered sequences

<400> 8
atggatgttt tgttgttgga aaaagctttg ttgggtttgt tcgccgcggc ggtgctggcc 60
atcgccgtcg ccaagctcac cggcaagcgc ttccgcctcc cccctggccc ctccggcgcc 120
cccatcgtcg gcaactggct gcaggtcggc gacgacctca accaccgcaa cctgatgggc 180
ctggccaagc ggttcggcga ggtgttcctc ctccgcatgg gcgtccgcaa cctggtggtc 240
gtctccagcc ccgagctcgc caaggaggtc ctccacaccc agggcgtcga gttcggctcc 300
cgcacccgca acgtcgtctt cgacatcttc accggcaagg gacaggacat ggtgttcacg 360
gtgtacggcg accactggcg caagatgcgg cggatcatga cggtgccctt cttcaccaac 420
aaggtggtgg cgcagaaccg cgtggggtgg gaggaggagg cccggctggt ggtggaggac 480
ctcaaggccg acccggcggc ggcgacggcg ggcgtggtgg tccgccgcag gctgcagctc 540

atgatgtaca acgacatgtt ccgcatcatg ttcgaccgcc ggttcgagag cgtggccgac 600
ccgctcttca accagctcaa ggcgctcaac gccgagcgca gcatcctctc ccagagcttc 660
gactacaact acggcgactt catccccgtc ctccgcccct tcctccgccg ctacctcaac 720
cgctgcacca acctcaagac caagcggatg aaggtgttcg aggaccactt cgtccagcag 780
cgcaaggagg cgttggagaa gacgggtgag atcaggtgcg ccatggacca catcctggaa 840
gccgaaagga agggcgagat caaccacgac aacgtcctct acatcgtcga gaacatcaac 900
gtcgcagcca tcgagacgac gctgtggtcg atcgagtggg gcctcgcgga gctggtgaac 960
cacccggaga tccagcagaa gctgcgcgag gagatcgtcg ccgttctggg cgccggcgtg 1020
gcggtgacgg agccggacct ggagcgcctc ccctacctgc agtccgtggt gaaggagacg 1080
ctccgcctcc gcatggcaat cccgctcctg gtgccgcaca tgaacctcag cgacgccaag 1140
ctcgccggct acgacatccc cgccgagtcc aagatcctcg tcaacgcctg gttcctcgcc 1200
aacgacccca agcggtgggt gcgcgccgat gagttcaggc cggagaggtt cctcgaggag 1260
gagaaggccg tcgaggccca cggcaacgat ttccggttcg tgcccttcgg cgtcggccgc 1320
cggagctgcc ccgggatcat cctcgcgctg cccatcatcg gcatcacgct cggacgcctg 1380
gtgcagaact tccagctgct gccgccgccg ggcaggaca agatcgacac caccgagaag 1440
cccgggcagt ttaccaacca gatcctcaag cacgccacca ttgtctgcaa gccactcgag 1500
gcttaa 1506

<210> 9
<211> 1506
<212> DNA
<213> Artificial Sequence

<220>
<223> Altered sequences

<400> 9
atggatgttt tgttgttgga aaaagctttg ttgggtttgt ttgctgctgc tgttttggct 60
attgctgttg ctaaattgac tggtaaaaga tttagattgc caccaggtcc atccggcgcc 120
cccatcgtcg gcaactggct gcaggtcggc gacgacctca accaccgcaa cctgatgggc 180
ctggccaagc ggttcggcga ggtgttcctc ctccgcatgg gcgtccgcaa cctggtggtc 240
gtctccagcc ccgagctcgc caaggaggtc ctccacaccc agggcgtcga gttcggctcc 300
cgcacccgca acgtcgtctt cgacatcttc accggcaagg gacaggacat ggtgttcacg 360
gtgtacggcg accactggcg caagatgcgg cggatcatga cggtgccctt cttcaccaac 420
aaggtggtgg cgcagaaccg cgtggggtgg gaggaggagg cccggctggt ggtggaggac 480
ctcaaggccg acccggcggc ggcgacggcg ggcgtggtgg tccgccgcag gctgcagctc 540
atgatgtaca acgacatgtt ccgcatcatg ttcgaccgcc ggttcgagag cgtggccgac 600
ccgctcttca accagctcaa ggcgctcaac gccgagcgca gcatcctctc ccagagcttc 660
gactacaact acggcgactt catccccgtc ctccgcccct tcctccgccg ctacctcaac 720
cgctgcacca acctcaagac caagcggatg aaggtgttcg aggaccactt cgtccagcag 780
cgcaaggagg cgttggagaa gacgggtgag atcaggtgcg ccatggacca catcctggaa 840
gccgaaagga agggcgagat caaccacgac aacgtcctct acatcgtcga gaacatcaac 900
gtcgcagcca tcgagacgac gctgtggtcg atcgagtggg gcctcgcgga gctggtgaac 960
cacccggaga tccagcagaa gctgcgcgag gagatcgtcg ccgttctggg cgccggcgtg 1020
gcggtgacgg agccggacct ggagcgcctc ccctacctgc agtccgtggt gaaggagacg 1080

ctccgcctcc gcatggcaat cccgctcctg gtgccgcaca tgaacctcag cgacgccaag 1140
ctcgccggct acgacatccc cgccgagtcc aagatcctcg tcaacgcctg gttcctcgcc 1200
aacgacccca agcggtgggt gcgcgccgat gagttcaggc cggagaggtt cctcgaggag 1260
gagaaggccg tcgaggccca cggcaacgat ttccggttcg tgcccttcgg cgtcggccgc 1320
cggagctgcc ccgggatcat cctcgcgctg cccatcatcg gcatcacgct cggacgcctg 1380
gtgcagaact tccagctgct gccgccgccg gggcaggaca agatcgacac caccgagaag 1440
cccgggcagt ttaccaacca gatcctcaag cacgccacca ttgtctgcaa gccactcgag 1500
gcttaa 1506

<210> 10
<211> 2181
<212> DNA
<213> Triticum aestivum

<400> 10
cgatccaccc cttggatcca ctctacccag ctcgctagcc agcggggtac atacacgcac 60
gcacgtacgc gcgtacgtac actcgcagag cttgcttcag ggaggccggc aatggaggtg 120
ggacgtgggg cggtggtggt gtcggcggtg gccgcgtaca tggcgtggtt ctggcggatg 180
tcccgcgggc tgcgcgggcc gcgggtttgg cccgtgctcg gcagcctgcc gggcctggtg 240
cagcacgccg aggacatgca cgagtggatc gccggcaacc tgcgccgcgc gggcggcacg 300
taccagacct gcatcttcgc cgtgcccggg gtggcgcgcc gcggcggcct ggtcaccgtc 360
acctgcgacc cgcgcaacct ggagcacgtc ctgaaggcgc gcttcgacaa ctaccccaag 420
ggcccettct ggcacggcgt cttccgggac ctgctcggcg acggcatctt caattccgac 480
ggcgacacct ggctcgcgca gcgcaagacg gccgcgctcg agttcaccac ccgcacgctc 540
cggacggcca tgtcccgctg ggtctcgcgc tccatccacg gccgcctcct gcccatcctg 600
gccgacgcgg ccaagggcaa ggcgcaggtg gatctccagg acctcctcct ccgcctcacc 660
ttcgacaaca tctgcggcct ggccttcggc aaggacccgg agacgctcgc ccagggcctg 720
ccggagaacg agttcgcctc cgcgttcgac cgcgccaccg aggccacgct caaccgcttc 780
atcttcccgg agttcctgtg gcgctgcaaa aagtggctgg gcctcggcat ggagaccacg 840
ctgaccagca gcatggccca cgtcgaccag tacctcgccg ccgtcatcaa gaagcgcaag 900
ctcgagctcg ccgccggcaa cggcaaatgc gacacggcgg cgacgcacga cgacctgctc 960
tcccggttca tgcggaaggg ttcctactcg gacgagtcgc tccagcacgt ggcgctcaac 1020
ttcatcctcg ccggccgcga cacctcctcc gtggcgctct cctggttctt ctggctcgtg 1080
tccacccacc ctgcggtgga gcgcaagatc gtgcgcgagc tctgctccgt tctcgccgcg 1140
tcacggggcg cccatgaccc ggcattgtgg ctggcggagc ccttcacctt cgaggagctc 1200
gaccgcctgg tctacctcaa ggcggcgctg tcggagaccc tccgcctcta cccctccgtc 1260
cccgaggact ccaagcacgt cgtcgcggac gactacctcc ccgacggcac cttcgtgccg 1320
gccgggtcgt cggtcaccta ctccatatac tcggcggggc gcatgaaggg ggtgtggggg 1380
gaggactgcc tcgagttccg gccggagcga tggctgtcgg ccgacggcac caagttcgag 1440
cagcacgact cgtacaagtt cgtggcgttc aacgccgggc cgagggtgtg cctgggcaag 1500
gacctagcct acctgcagat gaagaacatc gccgggagcg tgctgctccg gcaccgcctg 1560
accgtggcgc cgggccaccg cgtggagcag aagatgtcgc tcacgctctt catgaagggc 1620
gggctacgga tggaggtacg tccgcgcgac ctcgcccccg tcctcgacga gccctgcggc 1680
ctggacgccg gcgccgccac cgccgccgca gcaagtgcca cagcgccgtg cgcgtagaag 1740
acctggcacc ggcacgcgcc atgcatgatt cgtgcgtgct agctgttgaa gggacgccgg 1800

acattgaatg tgtagatagg gcagcagtgc aagaccgtaa gtaaaattga tgatgggttt 1860
ggtgacaaca ttgaagccac tcctttccag aatttacgac ccggatagga gaaacaggga 1920
aactttgcag atcacaacac aagatctagc cagccgggga tctgatctga tttgcgtctg 1980
ctcggagcac gggtgcatgg gagaccaagg aggaaaacaa aaaataacag aaacagagtg 2040
agcaatattt gtgattgtag ccacgggaaa gagagaggag taattagtaa ttcagatttg 2100
tttgcagtag ctcggtgttg gtgaccagat catagccaac taggctattc tattctattc 2160
tatttttgaa gatgattttt c 2181

<210> 11
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic primer

<400> 11
atatatggat ccatggaggt ggggacgtgg gcggtggtg 39

<210> 12
<211> 150
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic primer

<400> 12
atatatggat ccatggaagt tggtacttgg gctgttgttg tttctgctgt tgctgcttat 60
atggcttggt tttggagaat gtctagaggt ttgagaggtc caagagtttg gccagttttg 120
ggttctttgc caggcctggt gcagcacgcc 150

<210> 13
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic primer

<400> 13
tatatagaat tccttctacg cgcacggcgc tgtggcactt gc 42

<210> 14
<211> 1626

<212> DNA
<213> Artificial Sequence

<220>
<223> Altered sequences

<400> 14

```
atggaagttg gtacttgggc tgttgttgtt tctgctgttg ctgcttatat ggcttggttt      60
tggagaatgt ctagaggttt gagaggtcca agagtttggc cagttttggg ttctttgcca     120
ggcctggtgc agcacgccga ggacatgcac gagtggatcg ccggcaacct gcgccgcgcg     180
ggcggcacgt accagacctg catcttcgcc gtgcccgggg tggcgcgccg cggcggcctg     240
gtcaccgtca cctgcgaccc gcgcaacctg gagcacgtcc tgaaggcgcg cttcgacaac     300
taccccaagg gccccttctg gcacggcgtc ttccgggacc tgctcggcga cggcatcttc     360
aattccgacg gcgacacctg gctcgcgcag cgcaagacgg ccgcgctcga gttcaccacc     420
cgcacgctcc ggacggccat gtcccgctgg gtctcgcgct ccatccacgg ccgcctcctg     480
cccatcctgg ccgacgcggc caagggcaag gcgcaggtgg atctccagga cctcctcctc     540
cgcctcacct tcgacaacat ctgcggcctg gccttcggca aggacccgga gacgctcgcc     600
cagggcctgc cggagaacga gttcgcctcc gcgttcgacc gcgccaccga ggccacgctc     660
aaccgcttca tcttcccgga gttcctgtgg cgctgcaaaa agtggctggg cctcggcatg     720
gagaccacgc tgaccagcag catggcccac gtcgaccagt acctcgccgc cgtcatcaag     780
aagcgcaagc tcgagctcgc cgccggcaac ggcaaatgcg acacggcggc gacgcacgac     840
gacctgctct cccggttcat gcggaagggt tcctactcgg acgagtcgct ccagcacgtg     900
gcgctcaact tcatcctcgc cggccgcgac acctcctccg tggcgctctc ctggttcttc     960
tggctcgtgt ccacccaccc tgcggtggag cgcaagatcg tgcgcgagct ctgctccgtt    1020
ctcgccgcgt cacggggcgc ccatgacccg gcattgtggc tggcggagcc cttcaccttc    1080
gaggagctcg accgcctggt ctacctcaag gcggcgctgt cggagaccct ccgcctctac    1140
ccctccgtcc ccgaggactc caagcacgtc gtcgcggacg actacctccc cgacggcacc    1200
ttcgtgccgg ccgggtcgtc ggtcacctac tccatatact cggcggggcg catgaagggg    1260
gtgtgggggg aggactgcct cgagttccgg ccggagcgat ggctgtcggc cgacggcacc    1320
aagttcgagc agcacgactc gtacaagttc gtggcgttca acgccgggcc gagggtgtgc    1380
ctgggcaagg acctagccta cctgcagatg aagaacatcg ccgggagcgt gctgctccgg    1440
caccgcctga ccgtggcgcc gggccaccgc gtggagcaga agatgtcgct cacgctcttc    1500
atgaagggcg ggctacggat ggaggtacgt ccgcgcgacc tcgcccccgt cctcgacgag    1560
ccctgcggcc tggacgccgg cgccgccacc gccgccgcag caagtgccac agcgccgtgc    1620
gcgtag                                                               1626
```

<210> 15
<211> 501
<212> PRT
<213> Artificial Sequence

<220>
<223> Altered sequences

<400> 15

Met Asp Val Leu Leu Leu Glu Lys Ala Leu Leu Gly Leu Phe Ala Ala
1 5 10 15

Ala Val Leu Ala Ile Ala Val Ala Lys Leu Thr Gly Lys Arg Phe Arg
20 25 30

Leu Pro Pro Gly Pro Ser Gly Ala Pro Ile Val Gly Asn Trp Leu Gln
35 40 45

Val Gly Asp Asp Leu Asn His Arg Asn Leu Met Gly Leu Ala Lys Arg
50 55 60

Phe Gly Glu Val Phe Leu Leu Arg Met Gly Val Arg Asn Leu Val Val
65 70 75 80

Val Ser Ser Pro Glu Leu Ala Lys Glu Val Leu His Thr Gln Gly Val
85 90 95

Glu Phe Gly Ser Arg Thr Arg Asn Val Val Phe Asp Ile Phe Thr Gly
100 105 110

Lys Gly Gln Asp Met Val Phe Thr Val Tyr Gly Asp His Trp Arg Lys
115 120 125

Met Arg Arg Ile Met Thr Val Pro Phe Phe Thr Asn Lys Val Val Ala
130 135 140

Gln Asn Arg Val Gly Trp Glu Glu Glu Ala Arg Leu Val Val Glu Asp
145 150 155 160

Leu Lys Ala Asp Pro Ala Ala Ala Thr Ala Gly Val Val Val Arg Arg
165 170 175

Arg Leu Gln Leu Met Met Tyr Asn Asp Met Phe Arg Ile Met Phe Asp
180 185 190

Arg Arg Phe Glu Ser Val Ala Asp Pro Leu Phe Asn Gln Leu Lys Ala
195 200 205

Leu Asn Ala Glu Arg Ser Ile Leu Ser Gln Ser Phe Asp Tyr Asn Tyr
210 215 220

Gly Asp Phe Ile Pro Val Leu Arg Pro Phe Leu Arg Arg Tyr Leu Asn
225 230 235 240

Arg Cys Thr Asn Leu Lys Thr Lys Arg Met Lys Val Phe Glu Asp His
245 250 255

Phe Val Gln Gln Arg Lys Glu Ala Leu Glu Lys Thr Gly Glu Ile Arg
260 265 270

Cys Ala Met Asp His Ile Leu Glu Ala Glu Arg Lys Gly Glu Ile Asn
275 280 285

His Asp Asn Val Leu Tyr Ile Val Glu Asn Ile Asn Val Ala Ala Ile
290 295 300

Glu Thr Thr Leu Trp Ser Ile Glu Trp Gly Leu Ala Glu Leu Val Asn
305 310 315 320

His Pro Glu Ile Gln Gln Lys Leu Arg Glu Glu Ile Val Ala Val Leu
325 330 335

Gly Ala Gly Val Ala Val Thr Glu Pro Asp Leu Glu Arg Leu Pro Tyr

340 345 350

Leu Gln Ser Val Val Lys Glu Thr Leu Arg Leu Arg Met Ala Ile Pro
355 360 365
Leu Leu Val Pro His Met Asn Leu Ser Asp Ala Lys Leu Ala Gly Tyr
370 375 380
Asp Ile Pro Ala Glu Ser Lys Ile Leu Val Asn Ala Trp Phe Leu Ala
385 390 395 400
Asn Asp Pro Lys Arg Trp Val Arg Ala Asp Glu Phe Arg Pro Glu Arg
405 410 415
Phe Leu Glu Glu Glu Lys Ala Val Glu Ala His Gly Asn Asp Phe Arg
420 425 430
Phe Val Pro Phe Gly Val Gly Arg Arg Ser Cys Pro Gly Ile Ile Leu
435 440 445
Ala Leu Pro Ile Ile Gly Ile Thr Leu Gly Arg Leu Val Gln Asn Phe
450 455 460
Gln Leu Leu Pro Pro Pro Gly Gln Asp Lys Ile Asp Thr Thr Glu Lys
465 470 475 480
Pro Gly Gln Phe Thr Asn Gln Ile Leu Lys His Ala Thr Ile Val Cys
485 490 495
Lys Pro Leu Glu Ala
500

<210> 16
<211> 501
<212> PRT
<213> Artificial Sequence

<220>
<223> Altered sequences

<400> 16
Met Asp Val Leu Leu Leu Glu Lys Ala Leu Leu Gly Leu Phe Ala Ala
1 5 10 15
Ala Val Leu Ala Ile Ala Val Ala Lys Leu Thr Gly Lys Arg Phe Arg
20 25 30
Leu Pro Pro Gly Pro Ser Gly Ala Pro Ile Val Gly Asn Trp Leu Gln
35 40 45
Val Gly Asp Asp Leu Asn His Arg Asn Leu Met Gly Leu Ala Lys Arg
50 55 60
Phe Gly Glu Val Phe Leu Leu Arg Met Gly Val Arg Asn Leu Val Val
65 70 75 80
Val Ser Ser Pro Glu Leu Ala Lys Glu Val Leu His Thr Gln Gly Val
85 90 95
Glu Phe Gly Ser Arg Thr Arg Asn Val Val Phe Asp Ile Phe Thr Gly
100 105 110
Lys Gly Gln Asp Met Val Phe Thr Val Tyr Gly Asp His Trp Arg Lys

115 120 125
Met Arg Arg Ile Met Thr Val Pro Phe Phe Thr Asn Lys Val Val Ala
130 135 140
Gln Asn Arg Val Gly Trp Glu Glu Glu Ala Arg Leu Val Val Glu Asp
145 150 155 160
Leu Lys Ala Asp Pro Ala Ala Ala Thr Ala Gly Val Val Val Arg Arg
165 170 175
Arg Leu Gln Leu Met Met Tyr Asn Asp Met Phe Arg Ile Met Phe Asp
180 185 190
Arg Arg Phe Glu Ser Val Ala Asp Pro Leu Phe Asn Gln Leu Lys Ala
195 200 205
Leu Asn Ala Glu Arg Ser Ile Leu Ser Gln Ser Phe Asp Tyr Asn Tyr
210 215 220
Gly Asp Phe Ile Pro Val Leu Arg Pro Phe Leu Arg Arg Tyr Leu Asn
225 230 235 240
Arg Cys Thr Asn Leu Lys Thr Lys Arg Met Lys Val Phe Glu Asp His
245 250 255
Phe Val Gln Gln Arg Lys Glu Ala Leu Glu Lys Thr Gly Glu Ile Arg
260 265 270
Cys Ala Met Asp His Ile Leu Glu Ala Glu Arg Lys Gly Glu Ile Asn
275 280 285
His Asp Asn Val Leu Tyr Ile Val Glu Asn Ile Asn Val Ala Ala Ile
290 295 300
Glu Thr Thr Leu Trp Ser Ile Glu Trp Gly Leu Ala Glu Leu Val Asn
305 310 315 320
His Pro Glu Ile Gln Gln Lys Leu Arg Glu Glu Ile Val Ala Val Leu
325 330 335
Gly Ala Gly Val Ala Val Thr Glu Pro Asp Leu Glu Arg Leu Pro Tyr
340 345 350
Leu Gln Ser Val Val Lys Glu Thr Leu Arg Leu Arg Met Ala Ile Pro
355 360 365
Leu Leu Val Pro His Met Asn Leu Ser Asp Ala Lys Leu Ala Gly Tyr
370 375 380
Asp Ile Pro Ala Glu Ser Lys Ile Leu Val Asn Ala Trp Phe Leu Ala
385 390 395 400
Asn Asp Pro Lys Arg Trp Val Arg Ala Asp Glu Phe Arg Pro Glu Arg
405 410 415
Phe Leu Glu Glu Glu Lys Ala Val Glu Ala His Gly Asn Asp Phe Arg
420 425 430
Phe Val Pro Phe Gly Val Gly Arg Arg Ser Cys Pro Gly Ile Ile Leu
435 440 445
Ala Leu Pro Ile Ile Gly Ile Thr Leu Gly Arg Leu Val Gln Asn Phe
450 455 460
Gln Leu Leu Pro Pro Pro Gly Gln Asp Lys Ile Asp Thr Thr Glu Lys
465 470 475 480

Pro Gly Gln Phe Thr Asn Gln Ile Leu Lys His Ala Thr Ile Val Cys
485 490 495

Lys Pro Leu Glu Ala
500

<210> 17
<211> 501
<212> PRT
<213> Artificial Sequence

<220>
<223> Altered sequences

<400> 17
Met Asp Val Leu Leu Leu Glu Lys Ala Leu Leu Gly Leu Phe Ala Ala
1 5 10 15
Ala Val Leu Ala Ile Ala Val Ala Lys Leu Thr Gly Lys Arg Phe Arg
20 25 30
Leu Pro Pro Gly Pro Ser Gly Ala Pro Ile Val Gly Asn Trp Leu Gln
35 40 45
Val Gly Asp Asp Leu Asn His Arg Asn Leu Met Gly Leu Ala Lys Arg
50 55 60
Phe Gly Glu Val Phe Leu Leu Arg Met Gly Val Arg Asn Leu Val Val
65 70 75 80
Val Ser Ser Pro Glu Leu Ala Lys Glu Val Leu His Thr Gln Gly Val
85 90 95
Glu Phe Gly Ser Arg Thr Arg Asn Val Val Phe Asp Ile Phe Thr Gly
100 105 110
Lys Gly Gln Asp Met Val Phe Thr Val Tyr Gly Asp His Trp Arg Lys
115 120 125
Met Arg Arg Ile Met Thr Val Pro Phe Phe Thr Asn Lys Val Val Ala
130 135 140
Gln Asn Arg Val Gly Trp Glu Glu Glu Ala Arg Leu Val Val Glu Asp
145 150 155 160
Leu Lys Ala Asp Pro Ala Ala Ala Thr Ala Gly Val Val Val Arg Arg
165 170 175
Arg Leu Gln Leu Met Met Tyr Asn Asp Met Phe Arg Ile Met Phe Asp
180 185 190
Arg Arg Phe Glu Ser Val Ala Asp Pro Leu Phe Asn Gln Leu Lys Ala
195 200 205
Leu Asn Ala Glu Arg Ser Ile Leu Ser Gln Ser Phe Asp Tyr Asn Tyr
210 215 220
Gly Asp Phe Ile Pro Val Leu Arg Pro Phe Leu Arg Arg Tyr Leu Asn
225 230 235 240
Arg Cys Thr Asn Leu Lys Thr Lys Arg Met Lys Val Phe Glu Asp His

245 250 255
Phe Val Gln Gln Arg Lys Glu Ala Leu Glu Lys Thr Gly Glu Ile Arg
260 265 270
Cys Ala Met Asp His Ile Leu Glu Ala Glu Arg Lys Gly Glu Ile Asn
275 280 285
His Asp Asn Val Leu Tyr Ile Val Glu Asn Ile Asn Val Ala Ala Ile
290 295 300
Glu Thr Thr Leu Trp Ser Ile Glu Trp Gly Leu Ala Glu Leu Val Asn
305 310 315 320
His Pro Glu Ile Gln Gln Lys Leu Arg Glu Glu Ile Val Ala Val Leu
325 330 335
Gly Ala Gly Val Ala Val Thr Glu Pro Asp Leu Glu Arg Leu Pro Tyr
340 345 350
Leu Gln Ser Val Val Lys Glu Thr Leu Arg Leu Arg Met Ala Ile Pro
355 360 365
Leu Leu Val Pro His Met Asn Leu Ser Asp Ala Lys Leu Ala Gly Tyr
370 375 380
Asp Ile Pro Ala Glu Ser Lys Ile Leu Val Asn Ala Trp Phe Leu Ala
385 390 395 400
Asn Asp Pro Lys Arg Trp Val Arg Ala Asp Glu Phe Arg Pro Glu Arg
405 410 415
Phe Leu Glu Glu Glu Lys Ala Val Glu Ala His Gly Asn Asp Phe Arg
420 425 430
Phe Val Pro Phe Gly Val Gly Arg Arg Ser Cys Pro Gly Ile Ile Leu
435 440 445
Ala Leu Pro Ile Ile Gly Ile Thr Leu Gly Arg Leu Val Gln Asn Phe
450 455 460
Gln Leu Leu Pro Pro Pro Gly Gln Asp Lys Ile Asp Thr Thr Glu Lys
465 470 475 480
Pro Gly Gln Phe Thr Asn Gln Ile Leu Lys His Ala Thr Ile Val Cys
485 490 495
Lys Pro Leu Glu Ala
500

<210> 18
<211> 501
<212> PRT
<213> Artificial Sequence

<220>
<223> Altered sequences

<400> 18
Met Asp Val Leu Leu Leu Glu Lys Ala Leu Leu Gly Leu Phe Ala Ala
1 5 10 15

Ala Val Leu Ala Ile Ala Val Ala Lys Leu Thr Gly Lys Arg Phe Arg
20 25 30
Leu Pro Pro Gly Pro Ser Gly Ala Pro Ile Val Gly Asn Trp Leu Gln
35 40 45
Val Gly Asp Asp Leu Asn His Arg Asn Leu Met Gly Leu Ala Lys Arg
50 55 60
Phe Gly Glu Val Phe Leu Leu Arg Met Gly Val Arg Asn Leu Val Val
65 70 75 80
Val Ser Ser Pro Glu Leu Ala Lys Glu Val Leu His Thr Gln Gly Val
85 90 95
Glu Phe Gly Ser Arg Thr Arg Asn Val Val Phe Asp Ile Phe Thr Gly
100 105 110
Lys Gly Gln Asp Met Val Phe Thr Val Tyr Gly Asp His Trp Arg Lys
115 120 125
Met Arg Arg Ile Met Thr Val Pro Phe Phe Thr Asn Lys Val Val Ala
130 135 140
Gln Asn Arg Val Gly Trp Glu Glu Glu Ala Arg Leu Val Val Glu Asp
145 150 155 160
Leu Lys Ala Asp Pro Ala Ala Ala Thr Ala Gly Val Val Val Arg Arg
165 170 175
Arg Leu Gln Leu Met Met Tyr Asn Asp Met Phe Arg Ile Met Phe Asp
180 185 190
Arg Arg Phe Glu Ser Val Ala Asp Pro Leu Phe Asn Gln Leu Lys Ala
195 200 205
Leu Asn Ala Glu Arg Ser Ile Leu Ser Gln Ser Phe Asp Tyr Asn Tyr
210 215 220
Gly Asp Phe Ile Pro Val Leu Arg Pro Phe Leu Arg Arg Tyr Leu Asn
225 230 235 240
Arg Cys Thr Asn Leu Lys Thr Lys Arg Met Lys Val Phe Glu Asp His
245 250 255
Phe Val Gln Gln Arg Lys Glu Ala Leu Glu Lys Thr Gly Glu Ile Arg
260 265 270
Cys Ala Met Asp His Ile Leu Glu Ala Glu Arg Lys Gly Glu Ile Asn
275 280 285
His Asp Asn Val Leu Tyr Ile Val Glu Asn Ile Asn Val Ala Ala Ile
290 295 300
Glu Thr Thr Leu Trp Ser Ile Glu Trp Gly Leu Ala Glu Leu Val Asn
305 310 315 320
His Pro Glu Ile Gln Gln Lys Leu Arg Glu Glu Ile Val Ala Val Leu
325 330 335
Gly Ala Gly Val Ala Val Thr Glu Pro Asp Leu Glu Arg Leu Pro Tyr
340 345 350
Leu Gln Ser Val Val Lys Glu Thr Leu Arg Leu Arg Met Ala Ile Pro
355 360 365
Leu Leu Val Pro His Met Asn Leu Ser Asp Ala Lys Leu Ala Gly Tyr

370 375 380
Asp Ile Pro Ala Glu Ser Lys Ile Leu Val Asn Ala Trp Phe Leu Ala
385 390 395 400
Asn Asp Pro Lys Arg Trp Val Arg Ala Asp Glu Phe Arg Pro Glu Arg
405 410 415
Phe Leu Glu Glu Glu Lys Ala Val Glu Ala His Gly Asn Asp Phe Arg
420 425 430
Phe Val Pro Phe Gly Val Gly Arg Arg Ser Cys Pro Gly Ile Ile Leu
435 440 445
Ala Leu Pro Ile Ile Gly Ile Thr Leu Gly Arg Leu Val Gln Asn Phe
450 455 460
Gln Leu Leu Pro Pro Pro Gly Gln Asp Lys Ile Asp Thr Thr Glu Lys
465 470 475 480
Pro Gly Gln Phe Thr Asn Gln Ile Leu Lys His Ala Thr Ile Val Cys
485 490 495
Lys Pro Leu Glu Ala
500

<210> 19
<211> 541
<212> PRT
<213> Artificial Sequence

<220>
<223> Altered sequences

<400> 19
Met Glu Val Gly Thr Trp Ala Val Val Val Ser Ala Val Ala Ala Tyr
1 5 10 15
Met Ala Trp Phe Trp Arg Met Ser Arg Gly Leu Arg Gly Pro Arg Val
20 25 30
Trp Pro Val Leu Gly Ser Leu Pro Gly Leu Val Gln His Ala Glu Asp
35 40 45
Met His Glu Trp Ile Ala Gly Asn Leu Arg Arg Ala Gly Gly Thr Tyr
50 55 60
Gln Thr Cys Ile Phe Ala Val Pro Gly Val Ala Arg Arg Gly Gly Leu
65 70 75 80
Val Thr Val Thr Cys Asp Pro Arg Asn Leu Glu His Val Leu Lys Ala
85 90 95
Arg Phe Asp Asn Tyr Pro Lys Gly Pro Phe Trp His Gly Val Phe Arg
100 105 110
Asp Leu Leu Gly Asp Gly Ile Phe Asn Ser Asp Gly Asp Thr Trp Leu
115 120 125
Ala Gln Arg Lys Thr Ala Ala Leu Glu Phe Thr Thr Arg Thr Leu Arg
130 135 140

Thr Ala Met Ser Arg Trp Val Ser Arg Ser Ile His Gly Arg Leu Leu
145 150 155 160

Pro Ile Leu Ala Asp Ala Ala Lys Gly Lys Ala Gln Val Asp Leu Gln
165 170 175

Asp Leu Leu Leu Arg Leu Thr Phe Asp Asn Ile Cys Gly Leu Ala Phe
180 185 190

Gly Lys Asp Pro Glu Thr Leu Ala Gln Gly Leu Pro Glu Asn Glu Phe
195 200 205

Ala Ser Ala Phe Asp Arg Ala Thr Glu Ala Thr Leu Asn Arg Phe Ile
210 215 220

Phe Pro Glu Phe Leu Trp Arg Cys Lys Lys Trp Leu Gly Leu Gly Met
225 230 235 240

Glu Thr Thr Leu Thr Ser Ser Met Ala His Val Asp Gln Tyr Leu Ala
245 250 255

Ala Val Ile Lys Lys Arg Lys Leu Glu Leu Ala Ala Gly Asn Gly Lys
260 265 270

Cys Asp Thr Ala Ala Thr His Asp Asp Leu Leu Ser Arg Phe Met Arg
275 280 285

Lys Gly Ser Tyr Ser Asp Glu Ser Leu Gln His Val Ala Leu Asn Phe
290 295 300

Ile Leu Ala Gly Arg Asp Thr Ser Ser Val Ala Leu Ser Trp Phe Phe
305 310 315 320

Trp Leu Val Ser Thr His Pro Ala Val Glu Arg Lys Ile Val Arg Glu
325 330 335

Leu Cys Ser Val Leu Ala Ala Ser Arg Gly Ala His Asp Pro Ala Leu
340 345 350

Trp Leu Ala Glu Pro Phe Thr Phe Glu Glu Leu Asp Arg Leu Val Tyr
355 360 365

Leu Lys Ala Ala Leu Ser Glu Thr Leu Arg Leu Tyr Pro Ser Val Pro
370 375 380

Glu Asp Ser Lys His Val Val Ala Asp Asp Tyr Leu Pro Asp Gly Thr
385 390 395 400

Phe Val Pro Ala Gly Ser Ser Val Thr Tyr Ser Ile Tyr Ser Ala Gly
405 410 415

Arg Met Lys Gly Val Trp Gly Glu Asp Cys Leu Glu Phe Arg Pro Glu
420 425 430

Arg Trp Leu Ser Ala Asp Gly Thr Lys Phe Glu Gln His Asp Ser Tyr
435 440 445

Lys Phe Val Ala Phe Asn Ala Gly Pro Arg Val Cys Leu Gly Lys Asp
450 455 460

Leu Ala Tyr Leu Gln Met Lys Asn Ile Ala Gly Ser Val Leu Leu Arg
465 470 475 480

His Arg Leu Thr Val Ala Pro Gly His Arg Val Glu Gln Lys Met Ser
485 490 495

Leu Thr Leu Phe Met Lys Gly Gly Leu Arg Met Glu Val Arg Pro Arg

500 505 510
Asp Leu Ala Pro Val Leu Asp Glu Pro Cys Gly Leu Asp Ala Gly Ala
515 520 525
Ala Thr Ala Ala Ala Ala Ser Ala Thr Ala Pro Cys Ala
530 535 540

<210> 20
<211> 541
<212> PRT
<213> Artificial Sequence

<220>
<223> Altered sequences

<400> 20
Met Glu Val Gly Thr Trp Ala Val Val Val Ser Ala Val Ala Ala Tyr
1 5 10 15
Met Ala Trp Phe Trp Arg Met Ser Arg Gly Leu Arg Gly Pro Arg Val
20 25 30
Trp Pro Val Leu Gly Ser Leu Pro Gly Leu Val Gln His Ala Glu Asp
35 40 45
Met His Glu Trp Ile Ala Gly Asn Leu Arg Arg Ala Gly Gly Thr Tyr
50 55 60
Gln Thr Cys Ile Phe Ala Val Pro Gly Val Ala Arg Arg Gly Gly Leu
65 70 75 80
Val Thr Val Thr Cys Asp Pro Arg Asn Leu Glu His Val Leu Lys Ala
85 90 95
Arg Phe Asp Asn Tyr Pro Lys Gly Pro Phe Trp His Gly Val Phe Arg
100 105 110
Asp Leu Leu Gly Asp Gly Ile Phe Asn Ser Asp Gly Asp Thr Trp Leu
115 120 125
Ala Gln Arg Lys Thr Ala Ala Leu Glu Phe Thr Thr Arg Thr Leu Arg
130 135 140
Thr Ala Met Ser Arg Trp Val Ser Arg Ser Ile His Gly Arg Leu Leu
145 150 155 160
Pro Ile Leu Ala Asp Ala Ala Lys Gly Lys Ala Gln Val Asp Leu Gln
165 170 175
Asp Leu Leu Leu Arg Leu Thr Phe Asp Asn Ile Cys Gly Leu Ala Phe
180 185 190
Gly Lys Asp Pro Glu Thr Leu Ala Gln Gly Leu Pro Glu Asn Glu Phe
195 200 205
Ala Ser Ala Phe Asp Arg Ala Thr Glu Ala Thr Leu Asn Arg Phe Ile
210 215 220
Phe Pro Glu Phe Leu Trp Arg Cys Lys Lys Trp Leu Gly Leu Gly Met
225 230 235 240

```
Glu Thr Thr Leu Thr Ser Ser Met Ala His Val Asp Gln Tyr Leu Ala
                245                 250                 255
Ala Val Ile Lys Lys Arg Lys Leu Glu Leu Ala Ala Gly Asn Gly Lys
            260                 265                 270
Cys Asp Thr Ala Ala Thr His Asp Asp Leu Leu Ser Arg Phe Met Arg
        275                 280                 285
Lys Gly Ser Tyr Ser Asp Glu Ser Leu Gln His Val Ala Leu Asn Phe
    290                 295                 300
Ile Leu Ala Gly Arg Asp Thr Ser Ser Val Ala Leu Ser Trp Phe Phe
305                 310                 315                 320
Trp Leu Val Ser Thr His Pro Ala Val Glu Arg Lys Ile Val Arg Glu
                325                 330                 335
Leu Cys Ser Val Leu Ala Ala Ser Arg Gly Ala His Asp Pro Ala Leu
            340                 345                 350
Trp Leu Ala Glu Pro Phe Thr Phe Glu Glu Leu Asp Arg Leu Val Tyr
        355                 360                 365
Leu Lys Ala Ala Leu Ser Glu Thr Leu Arg Leu Tyr Pro Ser Val Pro
    370                 375                 380
Glu Asp Ser Lys His Val Val Ala Asp Asp Tyr Leu Pro Asp Gly Thr
385                 390                 395                 400
Phe Val Pro Ala Gly Ser Ser Val Thr Tyr Ser Ile Tyr Ser Ala Gly
                405                 410                 415
Arg Met Lys Gly Val Trp Gly Glu Asp Cys Leu Glu Phe Arg Pro Glu
            420                 425                 430
Arg Trp Leu Ser Ala Asp Gly Thr Lys Phe Glu Gln His Asp Ser Tyr
        435                 440                 445
Lys Phe Val Ala Phe Asn Ala Gly Pro Arg Val Cys Leu Gly Lys Asp
    450                 455                 460
Leu Ala Tyr Leu Gln Met Lys Asn Ile Ala Gly Ser Val Leu Leu Arg
465                 470                 475                 480
His Arg Leu Thr Val Ala Pro Gly His Arg Val Glu Gln Lys Met Ser
                485                 490                 495
Leu Thr Leu Phe Met Lys Gly Gly Leu Arg Met Glu Val Arg Pro Arg
            500                 505                 510
Asp Leu Ala Pro Val Leu Asp Glu Pro Cys Gly Leu Asp Ala Gly Ala
        515                 520                 525
Ala Thr Ala Ala Ala Ala Ser Ala Thr Ala Pro Cys Ala
    530                 535                 540
```